

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
17 octobre 2002 (17.10.2002)

PCT

(10) Numéro de publication internationale
WO 02/081741 A2

(51) Classification internationale des brevets⁷ : C12Q 1/68,
I/02, C12N 9/10, 15/12, A61K 38/45, 48/00, A01K
67/027, A61P 31/14

(21) Numéro de la demande internationale :
PCT/FR02/01169

(22) Date de dépôt international : 4 avril 2002 (04.04.2002)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
01/04598 4 avril 2001 (04.04.2001) FR

(71) Déposants (pour tous les États désignés sauf US) : IN-
STITUT PASTEUR [FR/FR]; 28 rue du Docteur Roux,
F-75724 Cedex 15 (FR). CENTRE NATIONAL DE LA
RECHERCHE SCIENTIFIQUE - CNRS [FR/FR]; 3 rue
Michel-Ange, F-75794 PARIS Cedex 16 (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : GUENET,
Jean-Louis [FR/FR]; 4 rue de l'Ecuyer, F-91160
LONGJUMEAU (FR). MASHIMO, Tomoji [JP/FR];
Maison du Japon, Cité Internationale de Paris, F-75690
PARIS Cedex 14 (FR). SIMON-CHAZOTTES, Do-
minique [FR/FR]; 51 rue du Général Leclerc, F-92130
ISSY-LES-MOULINEAUX (FR). MONTAGUTELLI,
Xavier [FR/FR]; 14 rue de Chaumont, Le Bois Dieu,
F-78125 HERMERAY (FR). FRENKIEL, Marie-Pas-
cale [FR/FR]; 15 rue Chaptal, F-92300 LEVALLOIS (FR).
DESPRES, Philippe [FR/FR]; 18 place de la Liberté,
F-92250 LA GARENNE-COLOMBES (FR). DEUBEL,

Vincent [FR/FR]; 29 boulevard du Lycée, F-92170
VANVES (FR). BONHOMME, François [FR/FR];
Université des Sciences et Techniques du Languedoc,
(Montpellier II)-UMR 5000, Bâtiment 13, Place Eugène
Bataillon, F-34095 MONTPELLIER Cedex 5 (FR). LU-
CAS, Marianne [FR/FR]; La Mare au Curé, INA P-G,
F-78850 THIERVAL-GRIGNON (FR).

(74) Mandataires : CABINET ORES etc.; 6 avenue de Mes-
sine, F-75008 Paris (FR).

(81) États désignés (national) : AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ,
DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN,
YU, ZA, ZM, ZW.

(84) États désignés (régional) : brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet
eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet
européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR,
IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN,
TD, TG).

Publiée :

— sans rapport de recherche internationale, sera republiée
dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abrévia-
tions, se référer aux "Notes explicatives relatives aux codes et
abréviations" figurant au début de chaque numéro ordinaire de
la Gazette du PCT.

WO 02/081741 A2

(54) Title: USE OF PRODUCTS OF GENES OF THE 2'-5' OLIGOADENYLATE SYNTHETASE FAMILY (OAS) FOR
SCREENING ANTIVIRAL AGENTS AND FOR DETECTING RESPONSIVENESS TO FLAVIVIRIDAE INFECTION

(54) Titre : UTILISATION DES PRODUITS DES GENES DE LA FAMILLE 2'-5' OLIGOADENYLATE SYNTHETASE (OAS)
POUR LE CRIBLAGE D'ANTIVIRAUX ET LA DETECTION DE LA SENSIBILITE A L'INFECTION PAR LES FLAVIVIRI-
DAE

(57) Abstract: The invention concerns the use of products of genes of the 2'-5' oligoadenylate synthetase family (OAS) for screen-
ing antiviral agents and for detecting responsiveness to infection by *Flavivirida*.

(57) Abrégé : Utilisation des produits des gènes de la famille 2'-5' oligoadénylate synthétase (OAS) pour le criblage d'antiviraux
et la détection de la sensibilité à l'infection par les *Flaviviridae*.

**UTILISATION DES PRODUITS DES GENES DE LA FAMILLE 2'-5'
OLIGOADENYLATE SYNTHETASE (OAS) POUR LE CRIBLAGE
D'ANTIVIRAUX ET LA DETECTION DE LA SENSIBILITE A
L'INFECTION PAR LES FLAVIVIRIDAE.**

La présente invention est relative à l'utilisation des produits des gènes de la famille 2'-5' oligoadénylate synthétase (OAS) pour le criblage d'antiviraux et la détection de la sensibilité à l'infection par les *Flaviviridae*.

La famille des *Flaviviridae* regroupe les virus du genre flavivirus responsables de pathologies humaines graves telles que la dengue, la fièvre jaune, les encéphalites transmises par les tiques, l'encéphalite japonaise, l'encéphalite à West-Nile et les virus des hépatites C et G. Si les flavivirus sont susceptibles de provoquer une morbidité et une mortalité importantes chez l'homme, l'infection est généralement asymptomatique et seule une fraction des individus infectés développent une maladie grave.

Les flavivirus sont des petits virus enveloppés. Leur génome est une molécule d'ARN monocaténaire de polarité positive d'environ 11 000 bases. L'ARN génomique est associé à plusieurs copies de la protéine de capsid C pour former la nucléocapside; elle est entourée d'une enveloppe virale constituée d'une double couche lipidique issue des membranes du réticulum endoplasmique (RE) dans lesquelles sont ancrées la protéine d'enveloppe E et la protéine de membrane M. L'ARN génomique des flavivirus contient un unique cadre de lecture ouvert d'environ 10500 nucléotides flanqué de deux courtes régions non codantes à ses extrémités 5' et 3'. Le génome est traduit en une polyprotéine d'environ 3400 acides aminés qui est le précurseur des protéines structurales C, prM (le précurseur intracellulaire de M) et E dans sa partie N-terminale et d'au moins sept protéines non structurales (NS) de NS1 à NS5 dans sa partie C-terminale.

De nombreux facteurs semblent intervenir dans la réaction d'un sujet à une infection virale : des facteurs viraux pourraient être responsables de la sévérité de la maladie, alors que la constitution génétique de l'hôte (humain ou mammifère non-humain) contribuerait à la sensibilité ou à la résistance à l'infection.

Des modèles murins ont permis d'établir l'existence d'une résistance

génétique à l'infection par les flavivirus. Il a été montré que certaines lignées de souris récemment dérivées de l'état sauvage et appartenant aux espèces *Mus musculus musculus* ou *Mus spretus* (Det. BSVR, BRVR, PRI, CASA/Rk et CAST/Ei) sont résistantes à l'infection par les flavivirus, alors que les lignées consanguines de laboratoire les plus courantes qui dérivent majoritairement de l'espèce *Mus musculus domesticus* n'y résistent pas (Sangster *et al.*, J.Virol., 1993, 67 : 340-347).

La résistance est contrôlée par au moins un locus autosomal dénommé *Flv*, localisé sur le chromosome 5, chez la souris et trois allèles *Flv^f*, *Flv^r* et *Flv^{mr}* confèrent respectivement la sensibilité, la résistance et la résistance intermédiaire à l'infection par les flavivirus. En utilisant une souche du flavivirus de l'encéphalite de la Vallée de Murray et des souris issues du croisement retour de la lignée de souris résistante C3H/RV avec les lignées de souris sensibles C3/He ou BALB/c, le locus *Flv* a été localisé dans une région de 0,9 cM du chromosome 5, chez la souris, entre les marqueurs *D5Mit68* et *D5Mit242* (G.R. Shellam *et al.*, Rev. Sci. Tech. Off. Epiz. 1998.17 :231-248.).

En dépit de l'existence de ces modèles murins de résistance génétique à l'infection par les flavivirus, aucun gène cellulaire de mammifère impliqué dans la résistance à l'infection par les *Flaviviridae* n'a encore été identifié, de manière certaine au niveau moléculaire. C'est pourquoi les Inventeurs se sont donnés pour but de pourvoir à des outils, aptes à permettre d'évaluer la sensibilité de l'hôte (humain ou mammifère non-humain) dans certaines infections virales particulièrement graves pour l'homme comme celles provoquées par les *Flaviviridae*.

En utilisant un modèle expérimental mettant en œuvre une nouvelle souche neurovirulente et neuroinvasive du virus West-Nile particulièrement virulente, dont le génome présente la séquence SEQ ID NO: 1, et dénommée ci-après souche IS-98-ST1, et des lignées de souris résistantes dérivant de géniteurs sauvages appartenant à l'espèce *Mus musculus musculus* ou *Mus spretus*, croisées en retour avec les lignées sensibles de laboratoire BALB/c ou C57BL/6, les Inventeurs ont précisé la localisation du locus *Flv* dans un intervalle de 0,2 à 0,4 cM du chromosome 5 de souris, comprenant la famille des gènes 2'-5' oligoadénylate synthétase (OAS) et ils ont montré que c'est le gène OAS qui confère la résistance à l'infection par les *Flaviviridae*. Trois

isoformes de l'OAS -L1, L2 et L3- ont été décrites chez la souris (Genbank Data Library n° X55982 [L1], X58077 [2] et M33863 [L3]). Leurs gènes présentent une forte homologie de séquence (plus de 80 %) (figure 1) avec celui de l'OAS p40/p46 humain (Genbank Data Library n° XM-007004 et XM-007005) (figure 2). On ne dispose que de peu d'informations sur les formes L1, L2 et L3 de l'OAS murin.

Le système 2-5A qui implique la famille des gènes OAS et la RNase L participe à la défense de l'hôte contre une infection virale (Castelli et al., *Biomed. And Pharmacother.*, 1998, **52**, 386-390). Le système 2-5A est une voie de dégradation des ARNs intracellulaire (pour revue, Rutherford *et al.*, *N.A.R.*, 1991, **9**, 1917-1924). L'expression des gènes OAS est induite par les interférons (IFN). Les IFNs appartiennent à un groupe de cytokines qui induisent un état anti-viral dans beaucoup de lignées cellulaires (Goodbourn *et al.*, *J. Gene Virol.*, 2000, **81**, 2341-2364). Les deux isoformes α/β composent les IFNs de type I : ils se fixent sur le même récepteur et provoquent des réponses similaires chez l'hôte. Les phagocytes mononucléés et les fibroblastes sont respectivement les principaux producteurs d'IFN- α et d'IFN- β mais certains types cellulaires peuvent produire les deux isoformes. L'infection virale d'une cellule hôte induit la production de l'IFN α/β qui par fixation sur les récepteurs kinases à tyrosine, présents à la surface cellulaire des cellules avoisinantes non infectées, va induire l'expression de plusieurs espèces protéiques qui seront déterminants dans les défenses anti-virales. Une perte de l'homéostasie calcique des cellules neuronales, notamment chez le rat, est aussi capable d'induire l'expression des gènes de la famille OAS (Paschen *et al.*, *Neuroscience Letters*, 1999, **263**, 109-112). Il a aussi été montré que la protéine C du virus de l'hépatite C active le promoteur de l'OAS p40/p46 chez l'homme (Naganuma *et al.*, *J. Virol.*, 2000, **74**, 8744-8750).

La molécule OAS produite sous une forme latente devient active en interagissant avec une molécule d'ARN bicaténaire. L'OAS active va alors polymériser l'ATP en oligomères ppp[A2'p]nA[2-5A] qui vont interagir de façon allostérique avec la forme normalement latente de la RNase L (84 kda) pour l'activer. Le site catalytique de la RNase L est localisé dans sa région carboxy-terminale et des séquence répétées de type ankyrine, une région d'homologie aux protéines kinases et un domaine qui est

prédit former un doigt de zinc sont aussi retrouvés. La ribonucléase active dégrade les ARN monocaténares en les clivant au niveau des motifs riches en UA et UU.

Au cours d'une infection virale, la RNase L clive les ARN cellulaires de types messagers et ribosomaux comme les ARN viraux monocaténares, bloquant ainsi la progression du cycle réplcatif du virus. La RNase L joue aussi un rôle dans la régulation de l'expression des gènes des facteurs pro-apoptotiques tels que *Bax* et les caspases (Castelli *et al.* Cell Death and Differentiation, 1998, 5, 313-320 ; Rush *et al.*, J. Interferon Cytokine Res., 2000, 20, 1091-1100).

Des facteurs capables de réguler la voie OAS ont été mis en évidence. Un inhibiteur de la RNase L murine a été identifié, RNase I. Cet inhibiteur a été montré moduler la répression du gène MyoD, un facteur de transcription spécifique aux cellules musculaires (Bisbal *et al.*, Mol. Cell Biol., 2000, 20, 4959-4969). La RNase I est aussi induite par le HIV de type 1 et participe ainsi à la diminution de la réponse antivirale de la cellule (Martinand *et al.*, J. Virol., 1999, 73, 290-296).

Trois domaines conservés ont été identifiés dans la 2'-5'- oligoadénylate synthétase : une boucle P suivie d'une séquence riche en asparagine dénommée boîte D et une région riche en lysine et en arginine dénommée région KR. Des mutants ponctuels de l'un de ces trois domaines de l'isoforme L1 murine (boucle P : K67R, K67M, G62A et G63A ; boîte D : D76N et D78N ; domaine KR : K200R et K200M) ont une activité enzymatique très réduite ou complètement abolie (Yamamoto *et al.*, J. Interferon Cytokine Res., 2000, 20 : 337-344).

Les Inventeurs ont mis en évidence des mutations dans la séquence nucléotidique des gènes OAS chez les souris sensibles à l'infection par les *Flaviviridae* ; ces mutations inactivent le gène OAS.

Ces éléments ont conduit les Inventeurs à mettre au point un modèle adapté au criblage de molécules aptes à stimuler spécifiquement l'activité des gènes OAS et/ou à la détection des sujets sensibles à une infection par les virus de la famille des *Flaviviridae* et de ce fait mauvais répondeurs à un traitement à l'interféron ; la mesure de l'activité 2'-5' OAS chez un individu ou un groupe d'individus représentatifs d'une population humaine permet d'évaluer le risque pour cet individu ou cette population à développer une forme grave de la maladie (forme aiguë mortelle pour les

arboviroses ou forme chronique pour l'hépatite C).

En conséquence, la présente invention a pour objet un procédé de criblage de molécules aptes à stimuler un gène de la famille OAS, caractérisé en ce qu'il comprend :

- la mise en culture de cellules, issues d'un mammifère non-humain *Flv^r/Flv^r* ou *Flv^r/Flv^s*,
- l'induction de l'expression des gènes OAS, par addition d'interféron α ou β ou par un stress calcique, notamment par addition d'EGTA, et
- la mise en contact desdites cellules avec la molécule à cribler, et
- la mesure de l'activité d'un gène OAS, par comparaison avec un échantillon témoin.

Ledit échantillon témoin est notamment constitué par des cellules de mammifères non-humain *Flv^s/Flv^s*.

Selon un mode de mise en œuvre avantageux dudit procédé, ledit mammifère non humain *Flv^r/Flv^r* est de préférence une lignée de souris résistantes dérivant de géniteurs sauvages appartenant à l'espèce *Mus musculus musculus*, ou *Mus spretus* ; elles peuvent de manière encore plus préférée être croisées en retour avec des lignées murines sensibles de laboratoire telles que BALB/c ou C57BL/6, de façon à obtenir des lignées dites congéniques.

Ces lignées pourront être utilisées dans de multiples expériences destinées à analyser les mécanismes de défense de la souris contre les infections virales à *flavivirus* et notamment pour l'analyse de la physiopathogénie de l'infection au niveau cellulaire. Elles pourront aussi servir à la mise au point de thérapeutiques d'un genre nouveau, lorsque les expériences en question nécessiteront l'utilisation de lots homogènes d'animaux ayant tous la même constitution génétique, afin de comparer leur comportement après infection ou non. Ces animaux, qui par définition seront histocompatibles entre eux et avec l'autre congénique permettront de réaliser, si besoin est, des transferts cellulaires.

Selon un autre mode de mise en œuvre avantageux dudit procédé, ledit gène OAS est un gène autologue.

Selon un autre mode de mise en œuvre avantageux dudit procédé,

ledit gène OAS est un gène hétérologue.

Selon un autre mode de mise en œuvre avantageux dudit procédé, l'activité du gène OAS peut-être mesurée par détermination :

- de la quantité de transcrits OAS par des techniques classiques qui en elles-mêmes sont connues de l'homme du métier (*Northern-blot*, RT-PCR...),
- de la quantité de protéines 2'-5'OAS produites, par des techniques classiques qui en elles-mêmes sont connues de l'homme du métier (ELISA, RIA, radioimmunoprécipitation, *Western-blot*...),
- du niveau d'activité 2'-5'OAS, par des techniques classiques qui en elles-mêmes sont connues de l'homme du métier, telles que celles décrites dans Witt et al., J. Interferon Res., 1993, 13, 17-23 ou
- des séquences des ARNm ou de l'ADN génomique, issues des gènes de la famille OAS : mise en évidence éventuelle de l'une des mutations précitées.

La présente invention a également pour objet l'utilisation des molécules d'acides nucléiques, sélectionnées dans le groupe constitué par :

- a) les molécules d'ADN génomique de mammifère (humain ou non-humain) correspondant à un *locus* de résistance à une infection par un *Flaviviridae*, lesquelles molécules sont constituées par un fragment de 0,2 à 0,4 cM incluant la famille des gènes OAS sauvages ou mutés,
- b) les ADNc desdites molécules en a),
- c) les protéines codées par lesdites molécules en a) ou en b),

pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

Au sens de la présente invention, on entend par "gène muté" un gène présentant au moins une substitution, une insertion ou une délétion d'au moins un nucléotide d'une région codante ou non-codante, par rapport à la séquence du gène sauvage.

Dans le cas où les gènes OAS sont des gènes sauvages, lesdits individus sont résistants à l'infection par un *Flaviviridae*, dans le cas où les gènes OAS

sont des gènes mutés. ils sont. de préférence, inactivés ; en conséquence, les individus porteurs desdites mutations sont sensibles à l'infection par lesdits virus.

Selon un mode de réalisation avantageux de ladite utilisation, lesdites molécules sont. de préférence. des séquences d'ADN génomique des gènes de la famille OAS. les ADNc desdites séquences et les protéines correspondantes.

La présente invention a également pour objet une molécule d'ADN génomique de mammifère (humain ou non-humain) correspondant à un *locus* de résistance à une infection par un *Flaviviridae*, caractérisée en ce qu'elle est constituée par un fragment de 0.2 à 0.4 cM incluant un gène OAS inactivé codant pour une protéine présentant au moins une mutation dans la séquence d'ADNc correspondante, laquelle mutation est sélectionnée dans le groupe constitué par une délétion des nucléotides en positions 100 à 102, 221 à 232, 577 à 578 et une insertion d'un codon stop en positions 807-809. en référence à la séquence de l'ADNc de l'isoforme L3 de souris (GENBANK M33863. figure 3A à 3C) ; l'introduction d'un codon stop prématuré (figure 3B) est responsable de la production d'une protéine 2'-5' OAS tronquée inactive, déletée des résidus C-terminaux (acides aminés 258 à 367, en référence à la séquence de l'isoforme L3 de souris GENBANK AAA37116, figure 4).

Les mutations telles que définies ci-dessus sont présentent uniquement chez les souris sensibles à l'infection par les *Flaviviridae* ; elles inactivent le gène OAS.

Selon un mode de réalisation avantageux de ladite molécule, elle comprend la séquence du marqueur *D5Mit368* du chromosome 5 de souris.

La présente invention a également pour objet l'utilisation d'un vecteur recombinant comprenant une molécule d'acide nucléique telle que définie ci-dessus pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

La présente invention a également pour objet l'utilisation de cellules contenant un vecteur recombinant comprenant une molécule d'acide nucléique telle que définie ci-dessus pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

La présente invention a également pour objet l'utilisation d'un

mammifère non-humain recombinant comprenant une molécule d'acide nucléique telle que définie ci-dessus pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

La présente invention a également pour objet une molécule d'acide nucléique constituée par un ADNc ou une séquence d'ADN génomique issue d'un gène de la famille OAS comme médicament destiné au traitement des infections par les virus de la famille des *Flaviviridae*.

La présente invention a également pour objet une protéine codée par un ADNc ou une séquence d'ADN génomique issue d'un gène de la famille OAS comme médicament destiné au traitement des infections par les virus de la famille des *Flaviviridae*.

La présente invention a également pour objet un procédé d'évaluation de la sensibilité d'un individu à l'infection par un virus de la famille des *Flaviviridae* et/ou de sa réponse à un traitement par l'interféron, caractérisé en ce qu'il comprend :

- la mise en culture de cellules à partir d'un échantillon de cellules d'un individu,
- l'induction de l'expression des gènes OAS par addition d'interféron α ou β ou par un stress calcique, notamment par addition d'EGTA, et
- la mesure de l'activité d'un des gènes OAS, par comparaison avec un échantillon de cellules, obtenues à partir d'un sujet témoin résistant à l'infection.

L'évaluation est notamment utile pour évaluer la réponse vaccinale, en vue de mettre au point des souches atténuées plus efficaces.

La présente invention a également pour objet des réactifs utiles pour mettre en œuvre l'un des procédés tels que définis ci-dessus : criblage, évaluation ou détection.

Parmi ces réactifs, on peut citer :

- les amorces de séquences SEQ ID NO :5 à SEQ ID NO :22 et
- les sondes correspondant respectivement aux positions 257-707 du transcrit de l'isoforme L3 murine et aux positions 1379-1874 du transcrit de l'isoforme L2 murine (SEQ ID NO : 31-32).

L'activité du gène OAS peut-être mesurée par les techniques telles que définies ci-dessus.

La présente invention a également pour objet des cellules eucaryotes transformées, caractérisées en ce qu'elles comprennent une molécule d'acide nucléique de mammifère (humain ou non-humain), telle que définie ci-dessus.

Lesdites cellules sont, de préférence, obtenues par recombinaison homologue à l'aide d'un vecteur approprié, conformément à la technique décrite dans la demande EP 0 419 621 ou le brevet US 5,792 632.

La présente invention a, également pour objet des mammifères non-humains transgéniques, caractérisés en ce qu'ils incluent au moins une copie d'une molécule d'acide nucléique, telle que définie ci-dessus.

De manière préférée, lesdits mammifères, notamment des souris, sont obtenus par injection *in ovo* (technique classique de Brister et al.) d'une molécule d'ADN selon l'invention contenant la région OAS provenant de souris sauvages, c'est-à-dire résistantes à l'infection.

La présente invention a également pour objet des mammifères non-humains recombinants, caractérisés en ce qu'ils sont porteurs d'au moins un allèle du gène OAS inactivé.

On obtient par exemple des souris *knock-out* pour l'ensemble des gènes OAS, par délétion desdits gènes par la technique cre-LoxP (voir Demande WO 97/06271)

Outre les dispositions qui précèdent, l'invention comprend encore d'autres dispositions, qui ressortiront de la description qui va suivre, qui se réfère à des exemples de mise en œuvre de l'objet de la présente invention, avec références aux dessins annexés dans lesquels :

- la figure 1 illustre la structure de la famille des gènes OAS murins,
- la figure 2 illustre la structure de la famille des gènes OAS humains
- la figure 3 (A, B et C) représente l'alignement de la séquence nucléotidique de l'ADNc des gènes OAS des souris sensibles à l'infection par les *Flaviviridae* (C57BL/6) et du gène OAS1 humain,
- la figure 4 représente l'alignement de la séquence en acides

aminés des isoformes L1, L2 et L3 de la 2'-5' OAS des souris sensibles à l'infection par les *Flaviviridae* (C57BL/6) et de l'isoforme p40/p46 humaine,

- les figures 5A à 5E représentent la comparaison de la séquence en acides aminés des protéines virales de la souche IS-98-ST1 (SEQ ID NO: 2) et de la souche New York 1999 (NY99 : Genbank AF196835),

- la figure 6 représente la cinétique de mortalité et la cinétique d'apparition des anticorps sériques spécifiques chez les souris *Flv^s/Flv^s* (BALB/c) infectées par la souche IS-98-ST1 du virus West-Nile (UFF : Unités Formant Foyer),

- la figure 7 représente la cinétique de propagation de la souche IS-98-ST1 dans le système nerveux central des souris sensibles (BALB/c),

- la figure 8 représente le protocole expérimental utilisé pour préciser la localisation du locus *Flv* sur le chromosome 5 de la souris et pour établir une lignée congénique BALB/c *Flv'*,

- la figure 9 représente la carte génétique du locus *Flv*, déterminée à partir des souris sensibles, issues du premier croisement en retour entre les lignées résistantes (MAI/Pas et MBT/Pas) et les lignées sensibles (C57BL/6 ou BALB/c). Les cases blanches représentent les allèles BALB/c ou C57BL/6 et les cases noires représentent les allèles MAI/Pas ou MBT/Pas,

- la figure 10 représente les haplotypes autour du locus *Flv* déterminant la sensibilité ou la résistance aux *Flaviviridae*, issus du premier croisement en retour (BC1) entre les lignées résistantes (MAI/Pas et MBT/Pas) et les lignées sensibles (C57BL/6 et BALB/c). Les lignes grisées représentent les allèles (BALB/c ou C57BL/6) et les lignes noires représentent les allèles MAI/Pas ou MBT/Pas,

- la figure 11 représente la carte génétique et la carte physique du locus *Flv* et la position du gène OAS dans ce locus,

- la figure 12 représente la distribution des allèles *Flv* chez les souris résistantes et sensibles issues du premier croisement en retour (BC1) entre les lignées résistantes (MAI/Pas et MBT/Pas) et les lignées sensibles (C57BL/6 et BALB/c),

- la figure 13 représente la carte physique du clone BAC RP23-39M18 de souris sensibles (C57BL/6) sur laquelle figure la position des marqueurs de type microsatellite et STS ainsi que la position des gènes OAS murins,

- la figure 14 représente la cinétique d'apparition des antigènes viraux dans les cellules Neuro 2a et les neurones primaires de souris sensibles (BALB/c) infectées par le virus West-Nile (souche IS-98-ST1),
- la figure 15 représente la mort par nécrose des cellules Neuro 2a infectées par le virus West-Nile (souche IS-98-ST1) : m.i., multiplicité d'infection,
- la figure 16 représente l'activité antivirale de l'IFN- α sur les cellules Neuro 2a infectées par la souche IS-98-ST1 du virus West-Nile : UFF, Unités Formant Foyer. et
- la figure 17 représente les amorces utilisées pour la détection des mutations dans les séquences codantes de l'isoforme L1 des gènes OAS des souris sensibles C57BL6.

Exemple 1 : Matériels et méthodes

1) Souris mises en oeuvre

- lignées de souris consanguines sensibles *Flv^s/Flv^s* C57BL6 et BALB/c (Janvier).
- lignées de souris résistantes (*Flv^r/Flv^r*), dérivées de souris sauvages appartenant à l'espèce *Mus musculus musculus*, MAI/Pas (capturées en Autriche dans la région d'Ilmitz) et MBT/Pas (capturées en Bulgarie dans la région de General Toshevo), *Mus spretus* (SEG/Pas et STF/Pas) et *Mus musculus domesticus* (WMP/Pas) (F. Bonhomme et al., 1996, *The laboratory mouse and its wild relatives*, « *Genetics variants and strains of the laboratory mouse* », S.R.M.F. Lyon, S.D.M. Brown, Oxford University Press, Oxford, 1577-1596).

2) Virus

a) isolement, amplification, purification et titration

Le virus West-Nile (WN) a été obtenu à partir du système nerveux central d'une cigogne manifestant des troubles neuropathologiques sévères en septembre 1998, à Eilat (Israël). L'infection de cellule VERO par cet isolat est cytolytique et l'immunofluorescence indirecte avec un ascite de souris immun spécifique du virus West-Nile est positive à 100 %. Le virus produit sur cellules VERO a été récolté et amplifié sur cellules de moustiques AP61 Desprès et al., *Virology*, 1993, **196**, 209-219).

Le passage 1 (ou P1) du virus WN sur cellules AP61 a été récolté 3 jours après l'infection : il possède un titre de $2,5 \times 10^8$ UFF/ml (Unité Formant Foyer) par la technique de titration sur cellules AP61 décrite dans Desprès et al. (*Virology*, 1993, 196, 209-219). L'inoculum P1 du virus WN sur cellules AP61 a été identifié comme la souche IS-98-ST1.

Un P2 a été obtenu, à partir de cellules AP61 infectées par la souche IS-98-ST1. P1 (titre: 6×10^7 UFF/ml). L'inoculum P2 de IS-98-ST1 est utilisé pour les épreuves de sensibilité à l'infection virale chez les souris adultes.

Un inoculum viral P3 de la souche IS-98-ST1 avec un titre de 5×10^7 UFF/ml a été produit sur cellules AP61. Une préparation virale hautement purifiée, obtenue selon le protocole de purification des flavivirions décrit dans Desprès et al. (*Virology*, 1993, 196, 209-219) a été obtenue à partir de 20 boîtes de 150 cm^2 de cellules AP61 récoltées 3 jours après l'infection par l'inoculum P3 du virus WN souche IS-98-ST1 (multiplicité d'infection de 0,4). La souche IS-98-ST1 purifiée en gradient de saccharose a un titre final de 2×10^{10} UFF/ml.

Les ARNs extraits de ce virus purifié sont utilisés pour amplifier les ADNc correspondant aux protéines virales C, prM et NS1.

b) Séquençage de l'ARN viral

Le génome viral a été extrait à partir du surnageant de culture des cellules VERO infectées de l'exemple 1 à l'aide du kit "QIAamp Viral RNA" (QIAGEN), en suivant les instructions du fabricant. 6 produits RT-PCR chevauchants ont été amplifiés à partir de ces ARNs en utilisant les amorces décrites par Lanciotti et al. (Science, 199, 286 :2333-). Les extrémités 5' et 3' du génome viral ont été amplifiées à l'aide d'amorces synthétisées d'après la séquence de la souche WN-NY99 (Genbank n° AF202541). Les ADNc obtenus ont été purifiés par chromatographie échangeuse d'ions et précipités dans 2 volumes d'isopropanol. Ensuite les ADNc ont été séquencés sur les deux brins en utilisant le kit "Taq Dye Deoxy Terminator Cycle Sequencing" (PERKIN ELMER CORP./APPLIED BIOSYSTEM) et les amorces espacées de 400 paires de bases sur le génome viral (Lanciotti et al., précité). Le séquençage a été réalisé avec 0,2 pmoles d'ADNc purifié et 30 pmoles d'amorces, en suivant le protocole recommandé par le fabricant. L'alignement des séquences est

réalisé à l'aide du logiciel CLUSTAL W.

La séquence génomique complète de la souche IS-98-ST1 du virus West-Nile correspond à la séquence SEQ ID NO :1.

L'alignement des séquences en acides aminés de la souche IS-98-ST1 (Seq ID NO: 2) et de la souche NY99, présenté à la figure 5, montre que la souche IS-98-ST1 isolée en Israël en 1998 et la souche NY-99 isolée à New York en 1999 sont très proches (divergence de moins de 0,2% au niveau des séquences en acides aminés).

Cependant, les différences observées dans la souche IS-98-ST1, respectivement dans les protéines E (A₅₁), NS1 (N₁₇), NS2A (R₁₆₄), NS2B (G₈₂, E₈₃), NS3 (P₄₉₆, E₅₂₁) et NS5 (S₅₄, N₂₈₀, A₃₇₂) sont potentiellement responsables de la neurovirulence et des propriétés neuroinvasives observées avec cette souche et peuvent servir de marqueur de virulence du virus West-Nile.

3) Cellules

Des neurones primaires, des cellules endothéliales et des astrocytes du système nerveux central de souris sensibles homozygotes pour l'allèle *Flv^r* [Swiss ou BALB/c (Janvier)] et de souris résistantes homozygotes ou hétérozygotes pour l'allèle *Flv^r* sont isolés d'embryons de 14 jours. Les neurones primaires sont mis en culture en milieu Neuobasal (Gibco BRL) supplémenté avec 20% de facteur de différenciation B27, 20 mM de glutamine et 40 mg/l de gentamycine comme antibiotique.

Des cellules de neuroblastome murin Neuro 2a (10⁴ cellules/cm²) sont cultivées en Labtek à 8 chambres (Nunc) en milieu MEM (GIBCO BRL) supplémenté avec 10% de sérum de veau fœtal (EUROBIO) et 1% d'acides aminés non essentiels (GIBCO BRL).

Des cellules d'hépatome humain HepG2 (ATCC n° HB8065) sont cultivées dans les conditions classiques telles que décrites dans Marianneau et al., *J. Virol.*, 1996, 77, 2547-2554.

4) Produits

L'interféron α (INF α A/D) est fourni par la société Biosource (PHC4044) et l'EGTA par SIGMA.

Exemple 2 : Les souris de lignées sauvages et de lignées consanguines de labora-

toire se différencient par leur sensibilité à l'infection par la souche neuroinvasive IS-98-ST1 du virus West-Nile.

1) Les lignées de souris sensibles.

Des souris BALB/c âgées de 6 semaines sont inoculées par la voie intrapéritonéale avec 100 UFF de la souche IS-98-ST1 virus West-Nile (UFF:DL50 = 10), préparée comme décrit à l'exemple 1.

Ces souris meurent à 100% avec un temps moyen de mortalité de 9 ± 2 jours (figure 6).

La cinétique de propagation de la souche IS-98-ST1 dans le système nerveux central de la souris sensible (BALB/c) a été analysée à partir des extraits de cerveau des souris infectés titrés sur cellules AP61, selon la technique décrite dans Després et al. (*J. Virol.*, 1998, 72, 823-829), précité. Les résultats montrent que le virus est détecté dans le système nerveux central (SNC) murin au 5^{ème} jour de l'infection et la production virale est maximale au 7^{ème} jour (Figure 7). Au 9^{ème} jour de l'infection, le virus n'est plus détecté dans le SNC murin (Figure 7).

La réplication du virus WN dans le SNC et les organes périphériques des souris infectées par la souche IS-98-ST1 est également détectée par immunohistologie, selon les protocoles classiques tels que décrits dans Després et al., 1998 (précité) et par hybridation *in situ*, selon les protocoles décrits à l'exemple 1.

Les anticorps sériques spécifiquement dirigés contre les protéines du virus WN sont titrés par ELISA selon le protocole décrit dans Després et al., 1993 (précité), en utilisant la souche IS-98-ST1 purifié sur gradient de saccharose telle que décrite à l'exemple 1, comme antigène. Les résultats montrent que les anticorps sériques apparaissent au 5^{ème} jour de l'infection et sont significativement détectés au 7^{ème} jour (figure 6).

2) Les souris de lignées résistantes.

Les souris des lignées SEG, WMP, STF et MAI qui dérivent de souris sauvages sont inoculées par la voie intrapéritonéale, avec 1000 UFF (100 DL50) de la souche IS-98-ST1 préparée selon le protocole décrit à l'exemple 1.

Contrairement aux souris de laboratoire qui sont sensibles à l'infection par la souche IS-98-ST1 et meurent en une dizaine de jours, ces souris

dérivant de souris sauvages sont résistantes à l'inoculation de la souche IS-98-ST1 et néanmoins permissives à la réplication de la souche IS-98-ST1. En effet, l'infection virale des souris dérivant de souris sauvages est asymptomatique bien que le virus se multiplie *in toto* comme le démontre la production d'anticorps sériques anti-WN à hauts titres ; en ELISA, les titres des sérums à la dilution 1:100 pour 10^6 UFF de virion purifié IS-98-ST1 sont supérieurs à 1 unité de D.O. à 450 nm.

Les souris résistantes à l'infection virale sont utilisées pour la production de sérums immuns spécifiquement dirigés contre les protéines de la souche IS-98-ST1 du virus WN. Trois semaines après inoculation du virus WN, les sérums prélevés de souris résistantes (0.045 ml par souris) sont mélangés, décomplémentés 30 min à 56°C puis dilués au 1:10 dans du DPBS* (V/V) supplémenté avec 0,2% (V/V) de Sérum Albumine bovine (Life Technologies) et 0,05% (P/V) d'azide de sodium. Les sérums dilués sont répartis en 0,2 ml et conservés à -20°C. Les sérums immuns dirigés contre la souche IS-98-ST1 sont utilisés aux dilutions finales de 1:500 pour l'immunofluorescence indirecte et au 1:1000 pour l'immunoprécipitation des protéines virales radiomarquées.

Exemple 3 : Localisation du locus *Flv* dans une région de 0,2 à 0,4 cM du chromosome 5 de souris contenant le gène OAS.

1) Méthodes

a) Modèle d'analyse de la résistance à l'infection par les *Flaviviridae* (figure 8)

Des souris mâles des lignées résistantes MAI/Pas et MBT/Pas sont croisées avec des souris femelles des lignées sensibles C57BL/6 et BALB/c. Les souris mâles de la génération F1 sont croisées en retour avec des souris femelles des lignées résistantes C57BL/6 et BALB/c pour donner une génération de souris de premier de premier croisement en retour (BC1).

Des souris BC1 âgées de 5 semaines sont inoculées par voie intrapéritonéale avec la souche IS-98-ST1, préparée selon le protocole décrit à l'exemple 1, dans les conditions décrites à l'exemple 2.

Les animaux sont observés tous les jours et les taux de mortalité et de survie sont déterminés 14 jours après l'infection.

b) génotypage des allèles *Flv*

Les allèles *Flv* des individus BC1 ont été cartographiés par PCR génomique à l'aide d'amorces spécifiques de 16 microsatellites du chromosome 5 (Catalogue Research Genetics) entourant le locus *Flv* (figures 9-11), selon les techniques courantes de biologie moléculaire en utilisant les protocoles standards tels que ceux décrits dans *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000. Wiley and son Inc. Library of Congress, USA).

2) Résultats

L'analyse de la distribution des allèles *Flv* chez les souris BC1 sensibles et résistantes à l'infection par la souche IS-98ST1 montre qu'un allèle *Flv*^r est suffisant pour conférer la résistance à l'infection (figure 12). Les résultats montrent également que dans ce modèle il existe une corrélation parfaite entre le phénotype résistant et la présence de l'allèle *Flv*^r et une corrélation presque parfaite entre le phénotype sensible et l'absence de l'allèle *Flv*^r (figure 12).

Le génotypage des allèles *Flv* montre que le locus *FLV* est localisé dans une région de 0,2 à 0,4 cM contenant le gène OAS1 (figures 9-11).

Exemple 4 : Les souris sensibles à l'infection par les *Flaviridae* possèdent un gène OAS muté.

1) Méthodes

La séquence génomique du gène OAS des souris sensibles (C57BL/6) a été déterminée à partir de la séquence du clone BAC RP23-39M18 (figure 11) selon les techniques courantes de biologie moléculaire en utilisant les protocoles standards tels que ceux décrits dans *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA).

Les ADNc de l'isoforme L1 du gène OAS murin ont été amplifiés par RT-PCR (kit Titan One Tube RT-PCR; Roche Biochemicals n° 1939 823) et séquencés par la technique de séquençage automatique (Automat) à l'aide des amorces suivantes (voir figure 17) :

Exon 1

F1 (SEQ ID NO:9) et R1 (SEQ ID NO:10) : produit d'amplification
de 258 pb

F1 (SEQ ID NO:9) et R2 (SEQ ID NO:11) : produit d'amplification
de 396 pb

Exon 2

F2 (SEQ ID NO:12) et R3 (SEQ ID NO:13) : produit d'amplification
de 297 pb

F2 (SEQ ID NO:12) et R4 (SEQ ID NO:14) : produit d'amplification
de 515 pb

Exon 3

F3 (SEQ ID NO:15) et R5 (SEQ ID NO:16) : produit d'amplification
de 288 pb

F3 (SEQ ID NO:15) et R6 (SEQ ID NO:17) : produit d'amplification
de 501 pb

F4 (SEQ ID NO:18) et R5 (SEQ ID NO:16) : produit d'amplification
de 112 pb

F4 (SEQ ID NO:18) et R6 (SEQ ID NO:17) : produit d'amplification
de 325 pb

Exon 4

F5 (SEQ ID NO:19) et R7 (SEQ ID NO:20) : produit d'amplification
de 244 pb

F5 (SEQ ID NO:19) et R8 (SEQ ID NO:21) : produit d'amplification
de 418 pb

F6 (SEQ ID NO:22) et R7 (SEQ ID NO:20) : produit d'amplification
de 156 pb

F6 (SEQ ID NO:22) et R8 (SEQ ID NO:21) : produit d'amplification
de 340 pb

2) Résultats

L'alignement de la séquence de l'ADNc des isoformes L1, L2 et L3 du gène OAS des souris sensibles (C57BL/6) avec la séquence de l'ADNc de l'isoforme L1 humaine (figure 3A, 3B et 3C) montre que l'isoforme L1 des souris

sensibles possède 3 délétions (nucléotides 100 à 102, 221-232 ; 576-577) et un codon stop prématuré en phase, situé en position (807-809).

L'alignement des séquences en acides aminés des isoformes L1, L2 et L3 du gène OAS des souris sensibles (C57BL/6) avec la séquence de l'isoforme L1 humaine (figure 4) montre que l'isoforme L1 des souris sensibles correspond à une isoforme L1 tronquée de la région C-terminale, laquelle région C-terminale comprend la séquence conservée RPVILDPADPT qui est impliquée dans l'activité enzymatique de la 2'-5' OAS. De plus, l'isoforme L1 des souris C57BL/6 ne possède pas les 4 premiers acides aminés GSSG du domaine GSSGKGTTLRGRSDADLVVF qui sont impliqués dans l'activité enzymatique de la 2'-5' OAS.

Exemple 5 : Modèle cellulaire d'étude de l'activité des gènes OAS

1) Infection de cultures primaires et de lignées cellulaires par la souche IS-98-ST1 du virus West-Nile

a) Matériels et méthodes

a₁) cultures primaires

Des neurones primaires et des astrocytes du SNC de souris sensibles homozygotes pour l'allèle *Flv^s* (souris Swiss, Janvier) sont préparés selon les protocoles classiques tels que décrits à l'exemple 1. Les cellules sont infectées par la souche IS-98-ST1 à une multiplicité d'infection de 20 UFF par cellule (m.i. de 20). L'effet cytopathique est observé en microscopie optique, la production virale est analysée par titration sur cellules AP61 comme décrit précédemment à l'exemple 1 et l'expression des antigènes viraux est analysée par radioimmunoprécipitation à l'aide d'un sérum immun de souris anti-West-Nile, selon les protocoles classiques tels que décrits dans Duarte Dos Santos et al. (*Virology*, 2000, 274, 292-308).

Les résultats montrent que 80% des neurones en culture produisent les antigènes viraux :

- leur profil en gel de polyacrylamide-SDS est présenté à la figure 14A.
- la production virale est de $[3,0 \pm 1,5] \times 10^6$ UFF/ml après 20 h d'infection et de $[7,0 \pm 0,5] \times 10^7$ UFF/ml à 40 h.
- les effets cytopathiques (ECPs) de type nécrotique sont observés

après 48 h d'infection virale.

En revanche, les astrocytes du SNC murin ne sont pas permissifs à la réplication du virus WN souche IS-98-ST1.

a₂) lignées cellulaires

Des cellules de neuroblastome murin Neuro 2a et des cellules d'hépatome humain HepG2, cultivées dans les conditions classiques telles que décrites dans Duarte Dos Santos et al. (précité) sont infectées à différentes multiplicité d'infection par le virus WN souche IS-98-ST1, préparé comme décrit à l'exemple 1. L'effet cytopathique est observé en microscopie optique, la production virale est analysée par titration sur cellules AP61 comme décrit précédemment à l'exemple 1 et l'expression des antigènes viraux est analysée par radioimmunoprécipitation à l'aide d'un sérum immun de souris anti-West-Nile, selon les protocoles classiques tels que décrits dans Duarte Dos Santos et al., précité.

Les résultats montrent que les cellules de neuroblastome murin Neuro 2a sont permissives à la réplication de la souche IS-98-ST1 du virus WN. Une m.i. de 4 est nécessaire pour infecter 80% des cellules Neuro 2a en monocouche. La production virale est de 10^7 UFF/ml (m.i. de 4) après 40 h d'infection et la mort cellulaire par nécrose est massive (Figure 15). La cinétique de production des antigènes majeurs prM, E et NS1 à partir de la polyprotéine virale présentée dans la figure 4B. montre que le demi-temps de formation de la glycoprotéine d'enveloppe E est d'environ 30 min. La protéine E de la souche IS-98-ST1 semble ne posséder qu'un seul résidu N-glycane (figure 14C).

Les résultats montrent également que les cellules d'hépatome humain HepG2 sont permissives à la réplication de la souche IS-98-ST1 du virus WN. A une m.i. de 10, la production virale est de $[2 \pm 1] \times 10^6$ UFF/ml après 48 h d'infection et les ECPs sont observés à partir de 72 h.

2) Analyse de l'effet de l'EGTA et de l'INF- α sur la réplication du virus West-Nile et détection de l'activité des gènes OAS

a) Matériels et cellules :

L'interféron alpha (INF- α A/D) est fourni par la société Biosource (PHC4044) et l'agent chélateur EGTA par Sigma..

Les cellules Neuro 2a sont cultivées en MEM supplémenté avec 10% de sérum de veau fœtal (SVF ; Eurobio) et 1% d'acides aminés non essentiels (Gibco BRL). Les neurones primaires isolés d'embryons de 14 jours de souris BALB/c sont mis en culture en milieu Neurobasal (Gibco BRL) supplémenté avec 20% de facteur de différenciation B27. 20 mM de glutamine et de la gentamycine comme antibiotique.

b) Protocoles d'infection et de traitement des cellules neuronales par l'INF- α et l'EGTA :

- Infection par la souche IS-98-ST1 du virus WN :

Les cellules Neuro 2a (10^4 cellules/cm²) cultivées en Labtek à 8 chambres (Nunc) sont infectées avec la souche virale avec une multiplicité d'infection de 4 (cycle unique de réplication) ou 0.1 (cycle répliatif biphasique) Unités Formant Foyer (UFF ; titre viral obtenu sur cellules de moustiques AP61) par cellule dans du MEM supplémenté avec 0,2% de sérum albumine pendant 90 min à 37°C. Les cellules infectées sont incubées avec du MEM à 2% SVF.

Les neurones primaires de souris BALB/c (# 2,5 10^5 cellules/cm²) déposés sur Labtek à 8 chambres (Nunc) sont infectés avec la souche virale avec une multiplicité d'infection de 20 UFF par cellule dans du milieu Neurobasal avec 2% SFV pendant 90 min à 37°C. Les cellules infectées sont incubées avec du milieu Neurobasal avec 2% SFV et 20% B27.

- Traitement des cellules infectées par l'INF- α

Les tapis cellulaires sont lavés 3 fois avec du milieu non supplémenté puis traités avec 20 UI/ml (cellules Neuro 2a) ou 100 UI/ml (neurones primaires) d'INF- α qui sont additionnés dans le milieu de culture.

- Traitement des cellules infectées par l'EGTA

Les monocouches cellulaires sont lavées 3 fois avec du PBS déplété en calcium et magnésium (Gibco BRL) puis incubées dans une solution EGTA à 1 mM dans du PBS pendant 150 min à 37°C. Les cellules Neuro 2a sont ensuite incubées dans du MEM supplémenté avec 10% SVF et les neurones primaires dans du milieu Neurobasal avec 20% de B27.

- Préparation des sondes OAS murines

Les neurones primaires de souris BALB/c traités par 20 UI/ml d'INF- α pendant 12 h sont lysés par la solution de lyse du kit ATLASTM Pure Total RNA Labeling System (Clontech, # PT3231-1) et l'ARN total extrait est précipité avec 3 volumes d'éthanol 95% en présence de 0,2 M LiCl pendant 18 h à -20°C. Un aliquot de l'ARN total (0,5 μ g) est utilisé comme matrice pour la synthèse par la technique RT-PCR (kit Titan One Tube RT-PCR; Roche Biochemicals n° 1939 823) des sondes OAS à l'aide des amorces OAS-1 à OAS-4 suivantes :

OAS-1 : GTCAGACGCTGACCTGGTG (SEQ ID NO: 5, positions 257-275 ; transcrit L3, [M33863])

OAS-2 : AGCTTCTCCTTACACAGTTGG (SEQ ID NO: 6, positions 686-707 ; transcrit L3, [M33863])

OAS-3 : ACAGTGCAGGTGTGTGAGC (SEQ ID NO: 7, positions 1379-1398 ; transcrit L2, [X58077])

OAS-4 : TCATGTCTCAGAAAGGAAAC (SEQ ID NO: 8, positions 1854-1874 ; transcrit L2, [X58077])

Le couple d'amorces OAS-1 et OAS-2 a été sélectionné afin d'amplifier une région de haute identité nucléotidique entre les 3 transcrits OAS (L1, L2 et L3) qui ont été identifiés chez la souris. Le couple d'amorces OAS-3 et OAS-4 a

été dessiné dans la région 5' non codante spécifique au transcrit L2. Les produits RT-PCR sont clonés à l'aide du TOPO TA cloning (Invitrogen).

Les ARN totaux de neurones primaires de souris BALB/c, non traités, traités avec l'INF- α ou l'EGTA ou infectés par le virus WN, en présence ou non d'INF- α ou d'EGTA, seront hybridés avec les sondes OAS en Northern blot.

- Activité anti-virale de l'INF- α sur le virus West-Nile

Après induction par l'INF- α , trois transcrits des gènes OAS, L1, L2, L3 sont observés (Rutherford *et al.*, 1991). Le transcrit L3 est détecté dès 4 h post-induction, L2 est observé à 12 h post-induction et enfin L1 est observé après 18 h post-induction.

La concentration d'INF- α 20 IU/ml pour les cellules Neuro 2a n'altère pas la viabilité cellulaire sur 24 h.

Les cellules Neuro 2a infectées par le virus WN sont incubées avec l'INF- α (voir §2) d'une part dès le début de l'infection ou d'autre part après 5 h ou 10 h d'infection. Ces temps d'incubation avec l'INF- α ont été sélectionnés en fonction d'une part des cinétiques de transcription des gènes OAS L1, L2 et L3 et d'autre part de la cinétique de réplication virale.

L'addition de l'INF- α dès le début de l'infection virale réduit de 85% (m.i. de 4) le nombre total de cellules Neuro 2a positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection. L'addition de l'INF- α à 5 h post-infection réduit de 65% (m.i. de 4) le nombre total de cellules Neuro 2a positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection. L'addition de l'INF- α à 10 h post-infection réduit de 50% (m.i. de 4) le nombre total de cellules Neuro 2a positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection.

- Activité anti-virale de l'EGTA sur le virus West-Nile

L'incubation de neurones primaires de rat dans un milieu sans calcium supplémenté avec 1 mM de l'agent chélateur EGTA pendant 150 min provoque une augmentation de 350% du niveau transcriptionnel du (ou des) membres de la famille OAS.

Le pré-traitement des neurones primaires de souris BALB/c avec

1 mM EGTA pendant 2 h 30 (voir §3) réduit de 50% le nombre total de cellules positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection. Le traitement des neurones primaires de souris BALB/c avec 1 mM EGTA pendant 2 h 30 après 2 h d'infection réduit de 50% le nombre total de cellules positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection. Le traitement des neurones primaires de souris BALB/c avec 1 mM EGTA pendant 2 h 30 après 6 h d'infection réduit de 20% le nombre total de cellules positives en antigènes viraux par immunofluorescence indirecte à 24 h post-infection.

REVENDICATIONS

1°) Procédé de criblage de molécules aptes à stimuler un gène de la famille OAS, caractérisé en ce qu'il comprend :

- la mise en culture de cellules, issues d'un mammifère non-humain *Flv'/Flv'* ou *Flv'/Flv^s*.

- l'induction de l'expression des gènes OAS par addition d'interféron α ou β ou par un stress calcique, notamment par addition d'EGTA,

- la mise en contact des cellules avec la molécule à cribler, et

- la mesure de l'activité d'un gène OAS, par comparaison avec un échantillon témoin.

2°) Procédé selon la revendication 1, caractérisé en ce que ledit mammifère non humain *Flv'/Flv'* est de préférence une lignée de souris résistantes dérivant de géniteurs sauvages appartenant à l'espèce *Mus musculus musculus* ou *Mus spretus*.

3°) Procédé selon la revendication 1 ou la revendication 2, caractérisé en ce que lesdites souris résistantes sont croisées en retour avec des lignées sensibles de laboratoire.

4°) Procédé selon la revendication 1 ou la revendication 2, caractérisé en ce que ledit gène OAS est sélectionné dans le groupe constitué par les gènes autologues et les gènes hétérologues.

5°) Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce que l'activité du gène OAS peut-être mesurée à l'aide de l'une des méthodes suivantes : détermination de la quantité de transcrits OAS, détermination de la quantité de protéines 2'-5'OAS produites, détermination du niveau d'activité 2'-5'OAS ou détermination de la séquence des ARNm ou de l'ADN génomique issues des gènes de la famille OAS.

6°) Molécule d'ADN génomique de mammifère (humain ou non-humain) correspondant à un *locus* de résistance à une infection par un *Flaviviridae*, caractérisée en ce qu'elle est constituée par un fragment de 0,2 à 0,4 cM incluant un gène OAS inactivé codant pour une protéine présentant au moins une mutation dans la séquence d'ADNc correspondante, laquelle mutation est sélectionnée dans le groupe

constitué par une délétion des nucléotides en positions 100 à 102, 221 à 232, 577 à 578 et une insertion d'un codon stop en positions 807-809, en référence à la séquence de l'ADNc de l'isoforme L3 de souris accessible sous le n° GENBANK M33863.

7°) Molécule d'acide nucléique selon la revendication 6, caractérisée en ce qu'elle comprend la séquence du marqueur *D5Mit368* du chromosome 5 de souris.

8°) Utilisation des molécules d'acides nucléiques sélectionnées dans le groupe constitué par:

a) les molécules d'ADN génomique de mammifère (humain ou non-humain) correspondant à un *locus* de résistance à une infection par un *Flaviviridae*, lesquelles molécules sont constituées par un fragment de 0,2 à 0,4 cM incluant la famille des gènes OAS sauvages ou mutés,

b) les ADNc desdites molécules en a),

c) les protéines codées par lesdites molécules en a) ou en b),

pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

9°) Utilisation selon la revendication 8, caractérisée en ce que lesdites molécules sont des séquences d'ADN génomique des gènes de la famille OAS, les ADNc desdites séquences et les protéines correspondantes

10°) Utilisation d'un vecteur recombinant comprenant une molécule d'acide nucléique selon la revendication 6 ou la revendication 7 ou une molécule d'acide nucléique telle que définie dans les revendications 8 ou 9, pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

11°) Utilisation de cellules contenant un vecteur recombinant comprenant une molécule d'acide nucléique selon la revendication 6 ou la revendication 7 ou une molécule d'acide nucléique telle que définie dans les revendications 8 ou 9, pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

12°) Utilisation d'un mammifère non-humain recombinant comprenant une molécule d'acide nucléique selon la revendication 6 ou la revendication 7 ou

une molécule d'acide nucléique telle que définie dans les revendications 8 ou 9, pour le criblage de molécules antivirales destinées au traitement des infections par les virus de la famille des *Flaviviridae*.

13°) Molécule d'acide nucléique constituée par un ADNc ou une séquence d'ADN génomique issue d'un gène de la famille OAS comme médicament destiné au traitement des infections par les virus de la famille des *Flaviviridae*.

14°) Protéine codée par un ADNc ou une séquence d'ADN génomique issue d'un gène de la famille OAS comme médicament destiné au traitement des infections par les virus de la famille des *Flaviviridae*.

15°) Procédé d'évaluation de la sensibilité d'un individu à l'infection par un virus de la famille des *Flaviviridae* et/ou de sa réponse à un traitement par l'interféron, caractérisé en ce qu'il comprend :

- la mise en culture de cellules à partir d'un échantillon de cellules d'un individu,
- l'induction de l'expression des gènes OAS par addition d'interféron α ou β ou par un stress calcique, notamment par addition d'EGTA, et
- la mesure de l'activité d'un des gènes OAS, par comparaison avec un échantillon de cellules, obtenues à partir d'un sujet témoin résistant à l'infection.

16°) Réactifs utiles pour mettre en œuvre les procédés selon les revendications 1 à 5 ou 15, caractérisés en ce qu'ils sont sélectionnés dans le groupe constitué par les amorces de séquences SEQ ID NO :5 à 22 ainsi que les sondes correspondant respectivement aux positions 257-707 du transcrit de l'isoforme L3 murine et aux positions 1379-1874 du transcrit de l'isoforme L2 murine (SEQ ID NO: 31-32).

17°) Cellules eucaryotes transformées, caractérisées en ce qu'elles comprennent une molécule d'acide nucléique de mammifère selon la revendication 6 ou la revendication 7 ou bien un ADNc ou une protéine telles que définies dans les revendications 8 ou 9.

18°) Mammifères non-humains transgéniques, caractérisés en ce qu'ils incluent au moins une copie d'une molécule d'acide nucléique, selon la revendication 6 ou la revendication 7 ou une molécule d'acide nucléique telle que définie

dans les revendications 8 ou 9.

19°) Mammifères non-humains recombinants, caractérisés en ce qu'ils sont porteurs d'au moins un allèle du gène OAS inactivé.

FIGURE 1

Les gènes de l'oligoadénylate synthétase (OAS) de la souris

No. gène	Genbank	ESTs de souris	ARNm	ORF (taille)	acides aminés	Exon	taille de l'Exon
3 OAS1(L3)	M33863	AW542285... (>42)	1.7Kpb ^b	1104 pb	367aa	7	183, 289, 185, 233, 157, 27, 39
4 OAS1(L2)	X58077	AI449562... (>41)	2.2Kpb ^b	1104 pb	367aa	7	183, 289, 185, 233, 157, 27, 39
6 OAS1(L1)	X55982	BF136899... (>2)	1.7Kb, 4Kb ^a	759 pb	252aa	4	180, 277, 185, 117
2 nouveau OAS1	-	AA794503+AA536884 (>4)	-	1086 pb	361aa	6	183, 288, 197, 233, 154, 30
p54OASL	AF068835	-	3084 pb	1422 pb	473aa	-	-

^a le numéro des gènes OAS correspond à celui indiqué dans le clone BAC 39M18

^b la taille de chaque ARNm est indiquée (Rutherford et al.)

2/23

La famille de l'oligoadénylate synthétase (OAS) humaine

No. gène	Genbank	ARNm	ORF (taille)	acides aminés	Exon	taille de l'Exon
1 OAS1	p40 X02874	1.6K pb	1095pb	364aa (346aa)	6	-
	p46 -	1.8K pb	1203pb	400aa (346aa)	6	-
3 OAS2	p68 M87284	2.8, 3.9, 4.5K pb	2064pb	687aa (683aa)	11	-
	p71 -	3.3K pb	2104pb	727aa (683aa)	11	-
2 OAS3	p100 AF063613	7Kb	3264pb	1087aa	16	-
	(AF251351)					
- OAS-RP	p30 AF063612	1.8K pb	765pb	254aa (219aa)	4	(203), 283, 170, 7
	p56 AF063611	2.0K pb	1542pb	513aa (346aa)	6	(203), 283, 178, 242, 148, 787,

FIGURE 2

4/23

L3_M33863_ ATGTGCTGCCAGCCTTTGATGTCCTGGGT CATGTTAATACTTCAGCAAGCCTGATCCCA
 L2_X58077_ ATGTGCTTCAGCCTTTGATGTCCTGGGT CATGGTAGTATCAATAAGAAGCCTAATCCCT
 L1_M55982_ ATGTGCTGCCAGCCTATGATTTACTGGATCATCTTAACATCCTCAAGAAGCCTAACCAAC
 hOAS1_X02874_ ATGTGCTGCCTGCCCTTTGATGCCCTGGGT CAGTTGACTGGCAGCTATAAACCTAACCCCC

 L3_M33863_ GAATCTATGCCATCCTCATCGAGGAATGTACCTCCCTGGGGAAGGATGGCGAGTTCTCTA
 L2_X58077_ TAATCTACACCATCCTCATCTGGGAATGTACCTCCCTGGGGAAGGATGGCGAGTTCTCTA
 L1_M55982_ AATTCTACGCCAATCTCATC-AGTGGCGTACCCGCC--GGGAAGGAGGGCAAGTTATCGA
 hOAS1_X02874_ AAATCTATGTCAAGCTCATCGAGGAGTGACCCGACCTGCAGAAAGAGGGCGAGTTCTCCA
 * ****
 L3_M33863_ CCTGCTTCACGGAGCTCCAGCGGAATTCCTGAAGCAGCGCCCAACCAAGCTGAAGAGTC
 L2_X58077_ CCTGCTTCACGGAGCTCCAGCGGAATTCCTGAAGCAGCGCCCAACCAAGCTGAAGAGTC
 L1_M55982_ TCTGCTTTATGGGGCTTCAGAAGTACTTCTGAACTGTCGCCCAACCAAGCTGAAGCGCC
 hOAS1_X02874_ CTTGCTTCACAGAACTACAGAGAGACTTCTGAAGCAGCGCCCAACCAAGCTGAAGCGCC

 L3_M33863_ TCATCCGCCTGGTCAAGCACTGGTACCAACTGTGTAAGGAGAAGCTGGGGAAGCCATTGC
 L2_X58077_ TCATCCGCCTGGTCAAACTGGTACCAACTGTGTAAGGAGAAGCTGGGGAAGCCACTGC
 L1_M55982_ TCATCCGCCTGGTCAAGCACTGGTACCAACTGTGTAAGGAGAAGCTGGGGAAGCCACTGC
 hOAS1_X02874_ TCATCCGCCTAGTCAAGCACTGGTACCAAAATTTGAAGAAGAAGCTTGGGAAGC---TGC

 L3_M33863_ CTCCACAGTACGCCCTAGAGTTGCTCACTGTCTTTGCCTGGGAACAAGGGAATGGATGTT
 L2_X58077_ CCCCACAGTATGCCCTGGAGCTACTCACTGTCTATGCCTGGGAACAAGGGAATGGATGTA
 L1_M55982_ CCCCACAGTATGCCCTGGAGCTGCTCACTCGATGCCTGGGAGTATGGGAGTCGAGTAA
 hOAS1_X02874_ CACCTCAGTATGCCCTGGAGCTCCTGACGGTCTATGCTTGGGAGCGAGGGAGCATGAAAA
 * **
 L3_M33863_ ATGAGTTCAACACAGCCAGGGCTTCCGGACCGTCTTGGAACTGGTCATCAATTATCAGC
 L2_X58077_ ATGAGTTCAACACAGCCAGGGCTTCCGGACCGTCTTGGAACTGGTCATCAATTATCAGC
 L1_M55982_ CTAATTTCAACACAGCCAGGGCTTCTGAACCGTCTTGGAACTGGTCACCAAGTACCAAC
 hOAS1_X02874_ CACATTTCAACACAGCCCAAGGATTTCCGGACCGTCTTGGAAATTAGTCATAAACTACCAAG

 L3_M33863_ ATCTTCGAATCTACTGGACAAAGTATTATGACTTTCAACACCAGGAGGTCTCCAAATACC
 L2_X58077_ ATCTTCGAATCTACTGGACAAAGTATTATGACTTTCAACACAAGGAGGTCTCCAAATACC
 L1_M55982_ AGCTTCGAATCTACTGGACAGTGTATTATGACTTTGACATCAAGAGGTCTCTGAATACC
 hOAS1_X02874_ AACTCTGCATCTACTGGACAAAGTATTATGACTTTAAAAACCCATTATTGAAAAGTACC
 * **
 L3_M33863_ TGCACAGACAGCTCAGAAAAGCCAGGCCTGTGATCCTGGACCCAGCTGACCCACAGGGA
 L2_X58077_ TGCACAGACAGCTCAGAAAAGCCAGGCCTGTGATCCTGGACCCAGCTGACCCACAGGGA
 L1_M55982_ TGCACCAACAGCTCAAAAAA-----
 hOAS1_X02874_ TGAGAAGGCAGCTCAGAAAACCCAGGCCTGTGATCCTGGACCCGCGGACCTACAGGAA
 **
 L3_M33863_ ATGTGGCCGGTGGGAACCCAGAGGGCTGGAGGCGGTTGGCTGAAGAGGCTGATGTGTGGC
 L2_X58077_ ATGTGGCTGGTGGGAACCCAGAGGGCTGGAGGCGGTTGGCTGAAGAGGCTGATGTGTGGC
 L1_M55982_ -----
 hOAS1_X02874_ ACTTGGGTGGTGGAGACCCAAAGGTTGGAGGCAGCTGGCACAAGAGGCTGAGGCTGGC

 L3_M33863_ TATGGTACCCATGTTTTATGAAAAAGGATGGTTCCTGAGTGAAGTCTCTGGGATGTGCCGA
 L2_X58077_ TGTGGTACCCATGTTTTATGAAAAATGATGGTTCCTGAGTGAAGTCTCTGGGATGTGCCGA
 L1_M55982_ -----
 hOAS1_X02874_ TGAATTACCCATGCTTTAAGAAATTGGGATGGGTCCCGAGTGAAGTCTCTGGATTCTGCTGG

 L3_M33863_ CGGTGGTTCCTGTACTTTTGTAGCAGGTAGAAGAGAAGTGGACATGTATCCTGCTGTGAG
 L2_X58077_ CGGTGGTTCCTGTACTTTTGTAGCAGGTGGAGGAGAAGTGGACATGTATCCTGCTGTGAG
 L1_M55982_ -----
 hOAS1_X02874_ TGAGACCTCCTGCTTCTCCCTGC-----CATTCATCCCTGCCCTCTCCATGAA

FIGURE 3B

5/23

L3_M33863_	CACAGCAGCACCTGCCAGGAGACTGCTGGTCAGGGGCATTTGCTGCTCTGCTGCAGGCC
L2_X58077_	CACAGCAGCACCTGCCAGGAGACTGCTGGTCAGGGGCATTTGCTGCTCTGCTGCAGGCC
L1_M55982_	-----
hOAS1_X02874_	GCTTGAGACATATAGCTGGAGACCATTCTTCCAAAGAACTTACCTCT-TGCCAAAGGCC
..	
L3_M33863_	CATGACCCAGTGAGGGAGGGCCCCACCTGGCATCAGACTCCGTGCTTCTGATGCCTGCCA
L2_X58077_	CATGACCCAGTGAGGCAGGGCCCCACCTGGCATCAGACTCCGTGCTTCTGATGCCTGCCA
L1_M55982_	-----
hOAS1_X02874_	ATTTA--TATTCATATAGTGACAGGCTGTGCTCATATTTTACAGTCATTTTGGTCACAA
L3_M33863_	GCCATGTTTGACTCCTGTCCAATCACAGCCAGCCTTCCTCAACAGATTCAGAAGGAGAGG
L2_X58077_	GCCATGTTTGACTCCTGTCCAATCACAGCCAGCCTTCCTCAACAGATTCAGAAGGAGAGG
L1_M55982_	-----
hOAS1_X02874_	TCGAGGGTT---TCTGGAATTTTACATCCCTTGTCAGAAATTCATTCCTTAAGAGTAA
L3_M33863_	AAAGAACACACGCTTGGTGTCCATCTGTCCACCTGTTGGAAGGTTCTGTCTGACAAAGTC
L2_X58077_	AAAGAACACACGCTTGGTGTCCATCTGTCCACCTGTTGGAAGGTTCTGTCTGACAAAGTC
L1_M55982_	-----
hOAS1_X02874_	TAATAAATAATCTCTAACACCAAAAA-----
L3_M33863_	TGATCAACAATAAACCACAGCAGGTGCC-GTCA
L2_X58077_	TGATCAACAATAAACCACAGCAGGTGCCGTCA
L1_M55982_	-----
hOAS1_X02874_	-----

FIGURE 3C

7/23

CI	C	MSKKPGGPGKSRVNMMLKRGMPRVLSLIGLKRAMLSLIDGKGPINF/LAL
FLA		MSKKPGGPGKSRVNMMLKRGMPRVLSLIGLKRAMLSLIDGKGPINF/LAL
CI		LAFFRTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSS
FLA		LAFFRTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSS
CI	p-m	KQKKRGGKTGIAMIGLIASVGAVTLSNFQGVMMTVNATDVTDOVITPT
FLA		KQKKRGGKTGIAMIGLIASVGAVTLSNFQGVMMTVNATDVTDOVITPT
CI		AAGKNLCIVRAMDVGYMCDOTITYECPVLSAGNDPEDI DCWCTKSAVYVR
FLA		AAGKNLCIVRAMDVGYMCDOTITYECPVLSAGNDPEDI DCWCTKSAVYVR
CI	M	YGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRYLVKTESW
FLA		YGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRYLVKTESW
CI	E	ILRNPGYALVPAVIGWMLGSNTMQRVVFVLLLVAPAYSFNCLGMSNRD
FLA		ILRNPGYALVPAVIGWMLGSNTMQRVVFVLLLVAPAYSFNCLGMSNRD
CI		FLEGVSGATWVDLVLEGDSVITMSKDKPTIDVKMMNMEANLAEVRSYC
FLA		FLEGVSGATWVDLVLEGDSVITMSKDKPTIDVKMMNMEANLAEVRSYC
CI		YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFGK
FLA		YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFGK
CI		GSIDTCAKACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
FLA		GSIDTCAKACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
CI		ATQAGRFSITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYYVMTVGTKT
FLA		ATQAGRFSITPAAPSYTLKLGEYGEVTVDCPRSGIDTNAYYVMTVGTKT
CI		FLVHREWFMDLNLWSSAGSTVWRNRETLMFEFPHATKQSVIALGSQEG
FLA		FLVHREWFMDLNLWSSAGSTVWRNRETLMFEFPHATKQSVIALGSQEG
CI		ALHQALAGAIPEVSSNTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFK
FLA		ALHQALAGAIPEVSSNTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFK
CI		FLGTPADTGHTVVLELQYTGTDGPCKVPISSVASLNDLTPVGRLVTVNP
FLA		FLGTPADTGHTVVLELQYTGTDGPCKVPISSVASLNDLTPVGRLVTVNP
CI		FVSVATANAKVLIELEPPFGDSYIIVGRGEQQINHHWHKSGSSIGKAFTT
FLA		FVSVATANAKVLIELEPPFGDSYIIVGRGEQQINHHWHKSGSSIGKAFTT

FIGURE 5A

8/23

CI	TLKGAQRLAALGDTAWDFGVS	GGVFTSVGKAVHQVF	GGAFRSLFGGMSWI
FLA	TLKGAQRLAALGDTAWDFGVS	GGVFTSVGKAVHQVF	GGAFRSLFGGMSWI
		
	NS1		
CI	TQGLLGALLLWMGINARDRS	IALTFLAVGGVLLFLSVNVHADT	GCAIDIS
FLA	TQGLLGALLLWMGINARDRS	IALTFLAVGGVLLFLSVNVHADT	GCAIDIS
		
CI	RQELRCGNGVFIHNDVEAWMDRY	KYYYPETPQGLAKIIQKAKKEGVCGLRS	
FLA	RQELRCGSGVFIHNDVEAWMDRY	KYYYPETPQGLAKIIQKAKKEGVCGLRS	
		
CI	VSRLHQMWAEAVKDELNTLLKENG	VDLSVVVEKQEGMYKSAPKRLTATTE	
FLA	VSRLHQMWAEAVKDELNTLLKENG	VDLSVVVEKQEGMYKSAPKRLTATTE	
		
CI	KLEIGWKA	WKSILFAPELANNTFVVDGPETKECPTQNR	AWNSLEVEDFG
FLA	KLEIGWKA	WKSILFAPELANNTFVVDGPETKECPTQNR	AWNSLEVEDFG
		
CI	FGLTSTRMFLKVRESNTTECD	SKIIGTAVKNNLAIHSDLSYWIESRLNDT	
FLA	FGLTSTRMFLKVRESNTTECD	SKIIGTAVKNNLAIHSDLSYWIESRLNDT	
		
CI	WKLERAVLGEVK	SCTWPETHLWGDGILES	DLIIPVTLAGPRSNHNRPG
FLA	WKLERAVLGEVK	SCTWPETHLWGDGILES	DLIIPVTLAGPRSNHNRPG
		
CI	YKTQNGP	WDEGRVEIDFDYCPGTTVTLSESCGHRGPATRTT	TESGKLIT
FLA	YKTQNGP	WDEGRVEIDFDYCPGTTVTLSESCGHRGPATRTT	TESGKLIT
		
	NS2A		
CI	DWCCRSCTLPPLRYQ	TDSGCWYGM	EIRPQRHDEKTLVQSQVNAYNADMID
FLA	DWCCRSCTLPPLRYQ	TDSGCWYGM	EIRPQRHDEKTLVQSQVNAYNADMID
		
CI	PFQGLLVVFLATQ	EVLRKRWTAKISMPAILIALLVLVFGGITYTDVLR	
FLA	PFQGLLVVFLATQ	EVLRKRWTAKISMPAILIALLVLVFGGITYTDVLR	
		
CI	VILVGAAFAESNS	GGDVVHLALMATFKIQPFVFMVASFLKARWTN	QENILL
FLA	VILVGAAFAESNS	GGDVVHLALMATFKIQPFVFMVASFLKARWTN	QENILL
		
CI	MLAAVFFQ	MAYHDARQILLWEIPDVLNSLAVAWMILRAITFTTTS	NNVVVP
FLA	MLAAVFFQ	MAYHDARQILLWEIPDVLNSLAVAWMILRAITFTTTS	NNVVVP
		
CI	LLALLTPRLRCLN	LDVYRILLMVGIGSLIREKRSAAAKKKGASLLCLAL	
FLA	LLALLTPGLRCLN	LDVYRILLMVGIGSLIREKRSAAAKKKGASLLCLAL	
		
	NS2B		
CI	ASTGLFNP	MILAAGLIACDPNRKRGWPATEVMTAVGLMFAIVGG	LAELDI
FLA	ASTGLFNP	MILAAGLIACDPNRKRGWPATEVMTAVGLMFAIVGG	LAELDI
		
CI	DSMAIPMTIAGLM	FAAFVISGKSTDMWIERTADISWESDAEITGSS	SERVD
FLA	DSMAIPMTIAGLM	FAAFVISGKSTDMWIERTADISWESDAEITGSS	SERVD
		
CI	VRLDDGENFQLMNDPGAPWKIWMRLRMVCLAISAYTPWAILPSVVGFWITL		

FIGURE 5B

9/23

FLA VRLDDDGNFQLMNDPGAPWKIWMMLRMVCLAISAYTPWAILPSVVGFWITL
.....
NS3

CI QYTKRGGVLWDTSPKEYKKGDTTTCGVYRIMTRGLLGSYQAGAGVMVEGV
FLA QYTKRGGVLWDTSPKEYKKGDTTTCGVYRIMTRGLLGSYQAGAGVMVEGV
.....

CI FHTLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEV
FLA FHTLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEV
.....

CI QMIVVEPGKNVKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGD
FLA QMIVVEPGKNVKNVQTKPGVFKTPEGEIGAVTLDFTGTSGSPIVDKNGD
.....

CI VIGLYGNGVIMPNGSYISAIVQGERMDEPIPAGEPEMLRKKQITVLDLH
FLA VIGLYGNGVIMPNGSYISAIVQGERMDEPIPAGEPEMLRKKQITVLDLH
.....

CI PGAGKTRRILPQIIKEAINRRLRTAVLAPTRVVAEMAEALRGLPIRYQT
FLA PGAGKTRRILPQIIKEAINRRLRTAVLAPTRVVAEMAEALRGLPIRYQT
.....

CI SAVPREHNGNEIVDMCHATLTHRLMSPHRVPNYNLFVMDEAHFTDPASI
FLA SAVPREHNGNEIVDMCHATLTHRLMSPHRVPNYNLFVMDEAHFTDPASI
.....

CI AARGYISTKVELGEAAAI FMTATPPGTSDFPESNSPISDLQTEIPORAW
FLA AARGYISTKVELGEAAAI FMTATPPGTSDFPESNSPISDLQTEIPORAW
.....

CI NSGYEWITEYTGKTWVFPVSVKMGNEIALCLQRAGKKVVQLNRKSYET EY
FLA NSGYEWITEYTGKTWVFPVSVKMGNEIALCLQRAGKKVVQLNRKSYET EY
.....

CI PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRAVILG
FLA PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRAVILG
.....

CI EPSAVTAASAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIM
FLA EPSAVTAASAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIM
.....

CI PDNINMPNGLIAQFYQPEREKVYTMGEYRLRGEERKNFLELLRTADLPV
FLA LDNINMPNGLIAQFYQPEREKVYTMGEYRLRGEERKNFLELLRTADLPV
.....

CI WLAYKVAAAGVSYHRRWCDFGPRNTILEDNNEVEVITKLGERKILRPR
FLA WLAYKVAAAGVSYHRRWCDFGPRNTILEDNNEVEVITKLGERKILRPR
.....

NS4A

CI WIDARVYSDHQALKAFKDFASGKRSQIGLIEVLGKMPEHFMGKTWEALDT
FLA WIDARVYSDHQALKAFKDFASGKRSQIGLIEVLGKMPEHFMGKTWEALDT
.....

CI MYVVATAEKGGRAHRMALEELPDALQTIALLSVMTMGVFFLLMQRK G
FLA MYVVATAEKGGRAHRMALEELPDALQTIALLSVMTMGVFFLLMQRK G
.....

CI IGKIGLGGAVLG VATFFCWMAEVPGT KIAGMLLLSLLLMI VLIPEPEKQR

FIGURE 5C

FLA IGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLSLLLMIVLPEPEKQR
.....
NS4B

CI SQTDNQLAVFLICVMTLVSAAVAAEMGWLDKTKSDISSLFGQRIEVKENF
FLA SQTDNQLAVFLICVMTLVSAAVAAEMGWLDKTKSDISSLFGQRIEVKENF
.....

CI SMGEFLDLRPAWLSYAVTTAVLTPLLKHLITSYINTSLTSINVQAS
FLA SMGEFLDLRPAWLSYAVTTAVLTPLLKHLITSYINTSLTSINVQAS
.....

CI ALFTLARGFPFVDVGVSALLAAGCWGQVTLTVTVTAATLLFCHYAYMVP
FLA ALFTLARGFPFVDVGVSALLAAGCWGQVTLTVTVTAATLLFCHYAYMVP
.....

CI GWQAEAMRSAQRRTAAGIMKNAVVDGIVATDVPELERTTPIMQKKVGQIM
FLA GWQAEAMRSAQRRTAAGIMKNAVVDGIVATDVPELERTTPIMQKKVGQIM
.....

CI LILVSLAAVVVNPSVKTVREAGILITAAAVTLWENGASSVWNATTAIGLC
FLA LILVSLAAVVVNPSVKTVREAGILITAAAVTLWENGASSVWNATTAIGLC
.....

NS5

CI HIMRGWLSCLSITWTLIKMEKPGKRGGAAGRTLGEVWKERLNQMTKE
FLA HIMRGWLSCLSITWTLIKMEKPGKRGGAAGRTLGEVWKERLNQMTKE
.....

CI EFTRYRKEAIIIEVDRSAAKHARKEGNVTGGHVSVRGTAKLRWLVERRFLE
FLA EFTRYRKEAIIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVERRFLE
.....

CI PVGKVIDLGCGRGGWCYYMATQKRVQEVRYTKGGPGHEEPQLVQSYGWN
FLA PVGKVIDLGCGRGGWCYYMATQKRVQEVRYTKGGPGHEEPQLVQSYGWN
.....

CI IVTMKSGVDVFYRPECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWL
FLA IVTMKSGVDVFYRPECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWL
.....

CI HRGPREFCVKVLCPYMPKVIKMEQLLQRRYGGGLVRNPLSRNSTHEMYWV
FLA HRGPREFCVKVLCPYMPKVIKMEQLLQRRYGGGLVRNPLSRNSTHEMYWV
.....

CI SRASGNVVHVSVMNTSQVLLGRMEKRTWKGPQYEEVDNLGSGTRAVGKPLL
FLA SRASGNVVHVSVMNTSQVLLGRMEKRTWKGPQYEEVDNLGSGTRAVGKPLL
.....

CI NSDTSKINNRIERLRREYSSTWHHDENHPYRTWNYHGSYDVKPTGSASSL
FLA NSDTSKINNRIERLRREYSSTWHHDENHPYRTWNYHGSYDVKPTGSASSL
.....

CI VNGVVRLLSKPWDITITNVTTMAMTDTTPFGQQRVFKEKVDTKAPEPPEGA
FLA VNGVVRLLSKPWDITITNVTTMAMTDTTPFGQQRVFKEKVDTKAPEPPEGV
.....

CI KYVLNETTNWLWAFIAREKRPRMCSREEFIRKVNSNAALGAMFEEQNQWR
FLA KYVLNETTNWLWAFIAREKRPRMCSREEFIRKVNSNAALGAMFEEQNQWR
.....

CI SAREAVEDPKFWEMVDEEREHLRGECHTCIYNNMGKREKKPGEFGKAKG

FIGURE 5D

11/23

FLA	SAREAVEDPKFWEMVDEEREAREHLRGECHTCIYNMMGKREKKPGEFGKAKG
CI FLA	SRAIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYYLR SRAIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYYLR
CI FLA	EVGTRPGGKIYADDTAGWDTRITRADLENEAKVLELLDGEHRLARAIIE EVGTRPGGKIYADDTAGWDTRITRADLENEAKVLELLDGEHRLARAIIE
CI FLA	LTYRHKVVKVMRPAADGRTVMDVISREDQRGSGQVVTYALNTFTNLAVQL LTYRHKVVKVMRPAADGRTVMDVISREDQRGSGQVVTYALNTFTNLAVQL
CI FLA	VRMMEGEGVIGPDDVEKLTGKGPKVRTWLFENGEERLSRMAVSGDDCVV VRMMEGEGVIGPDDVEKLTGKGPKVRTWLFENGEERLSRMAVSGDDCVV
CI FLA	KPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWYDWQQVPFCSNHFTELI KPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWYDWQQVPFCSNHFTELI
CI FLA	MKDGRITLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYAQMWLLLYF MKDGRITLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYAQMWLLLYF
CI FLA	HRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW HRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
CI FLA	IEENEWMEDKTPVEKWSQVPYSGKREDIWCGLIGTRAPATWAENIQVAI IEENEWMEDKTPVEKWSQVPYSGKREDIWCGLIGTRAPATWAENIQVAI
CI FLA	NQVRAIIGDEKYVDYMSSLKRYEDTTLVEDTVL NQVRAIIGDEKYVDYMSSLKRYEDTTLVEDTVL

FIGURE 5E

12/23

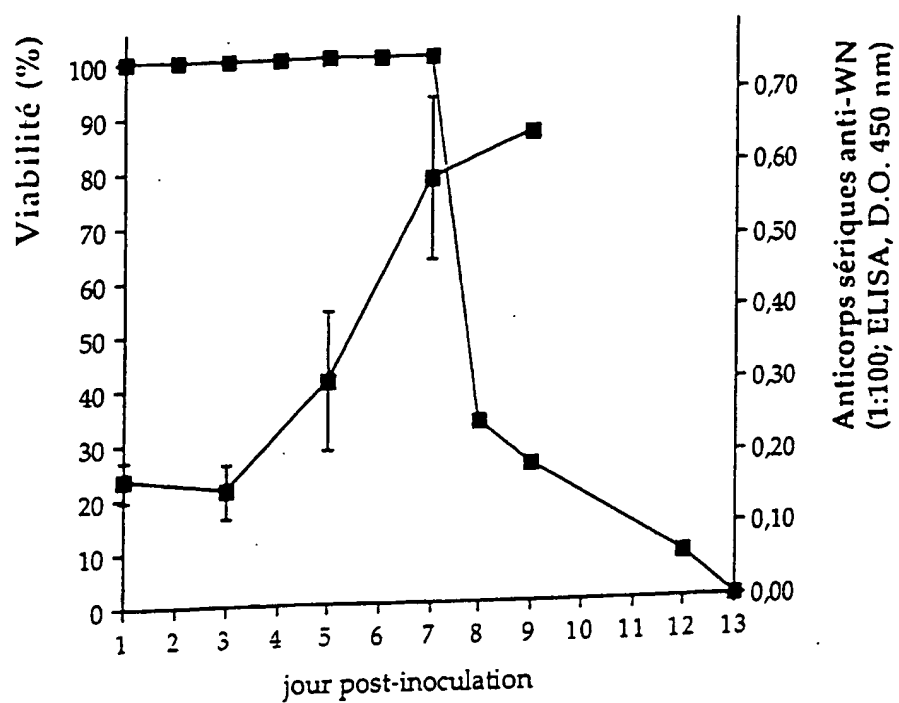


FIGURE 6

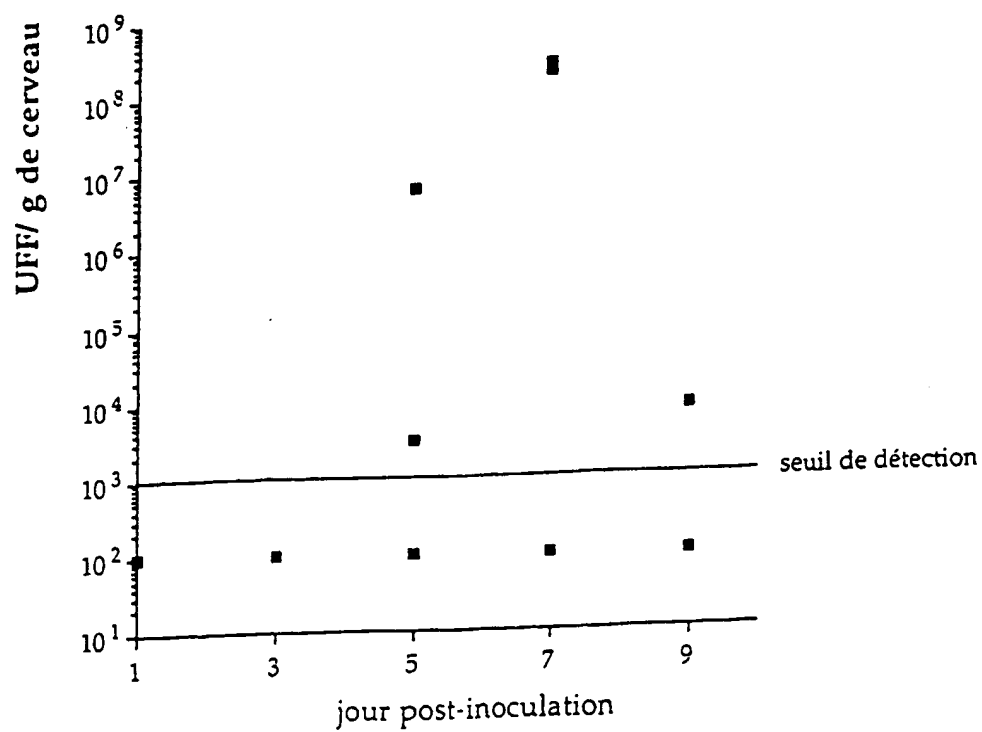


FIGURE 7

13/23

Souris

Lignées consanguines de laboratoire :

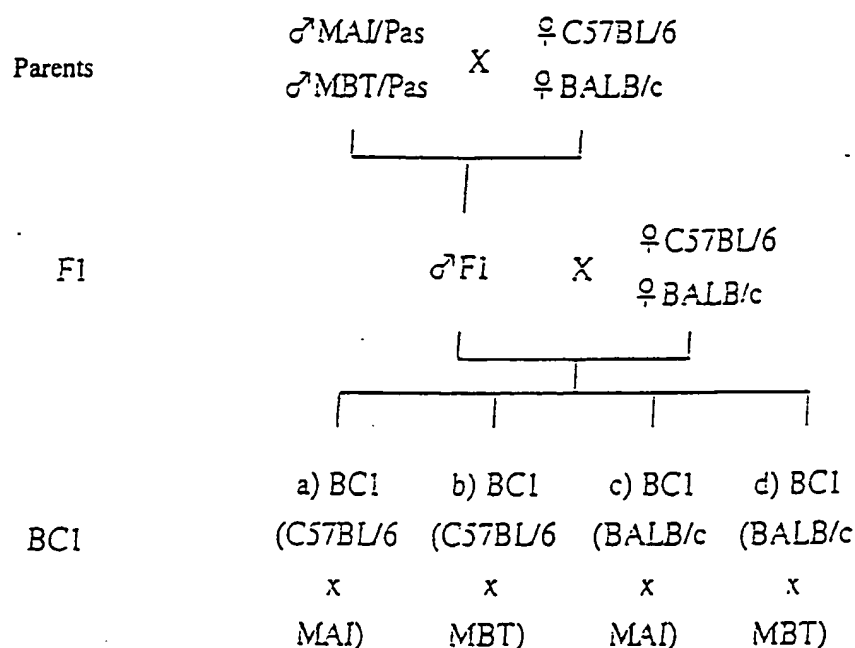
BALB/c, C57BL/6, DDK, 129, C3H et DBA/1

→ sensibles à l'infection par le virus WN

Souris sauvages :

SEG/Pas (*Mus spretus*), MAI/Pas, MBT/Pas (*Mus m. musculus*)

→ résistantes à l'infection par le virus WN

Génération de souris de premier croisement en retour (BC1)**Virus**

Injection du virus West Nile (WN)

Souris âgées de 5 semaines

Observation des souris pendant les 14 jours suivant l'infection

Génotypage des allèles *Flv*marqueurs flanquant le locus *Flv* sur le chromosome 5 de la souris

FIGURE 8

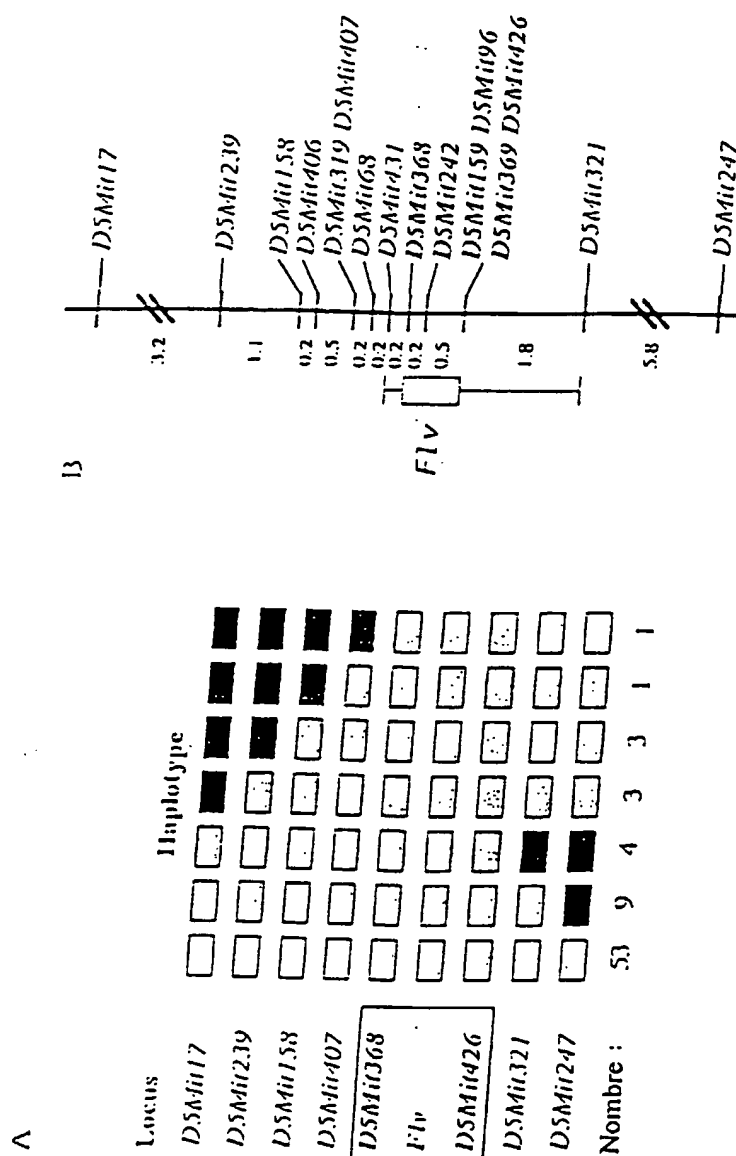


FIGURE 9

15/23

Parents de la génération de premier croisement en retour (BC1)

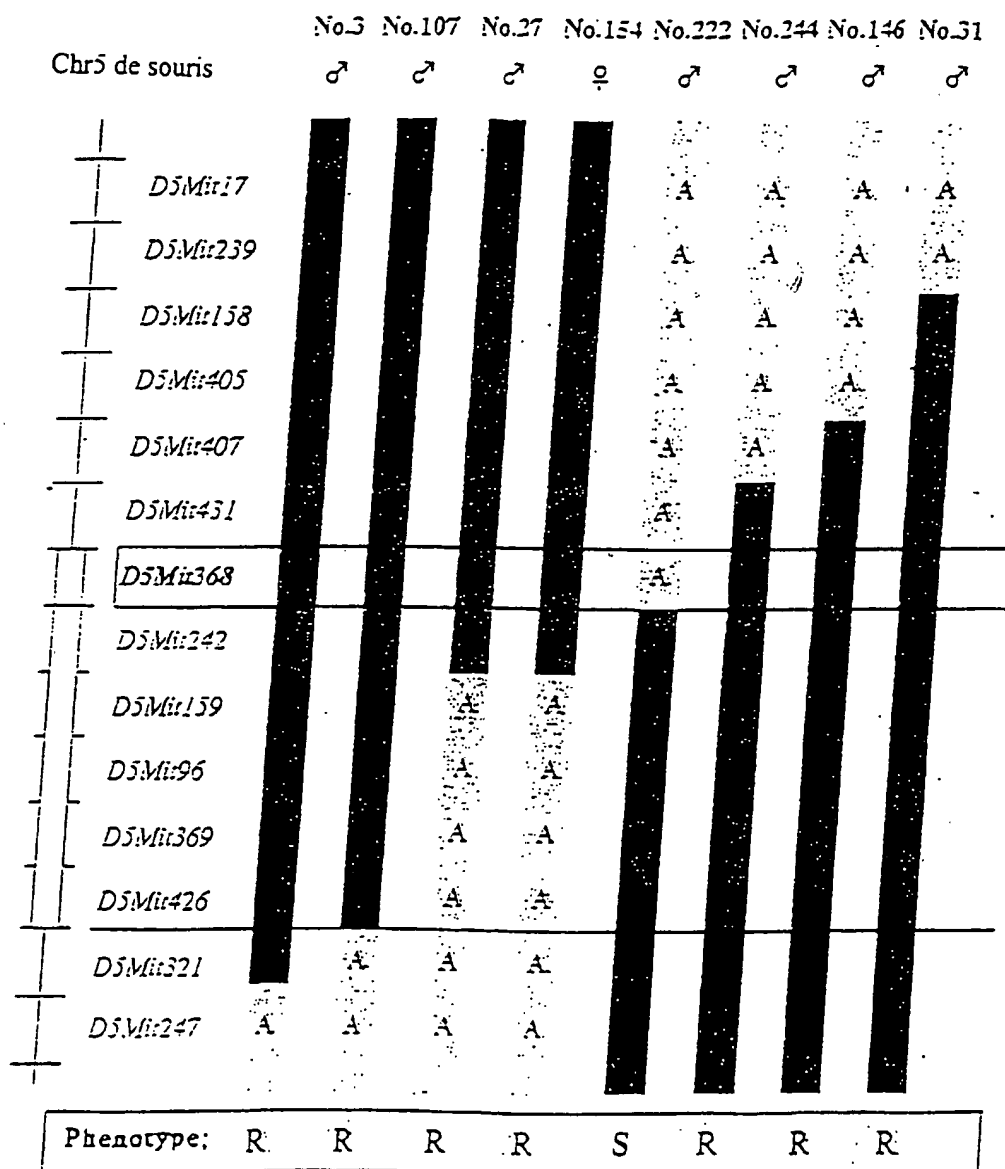


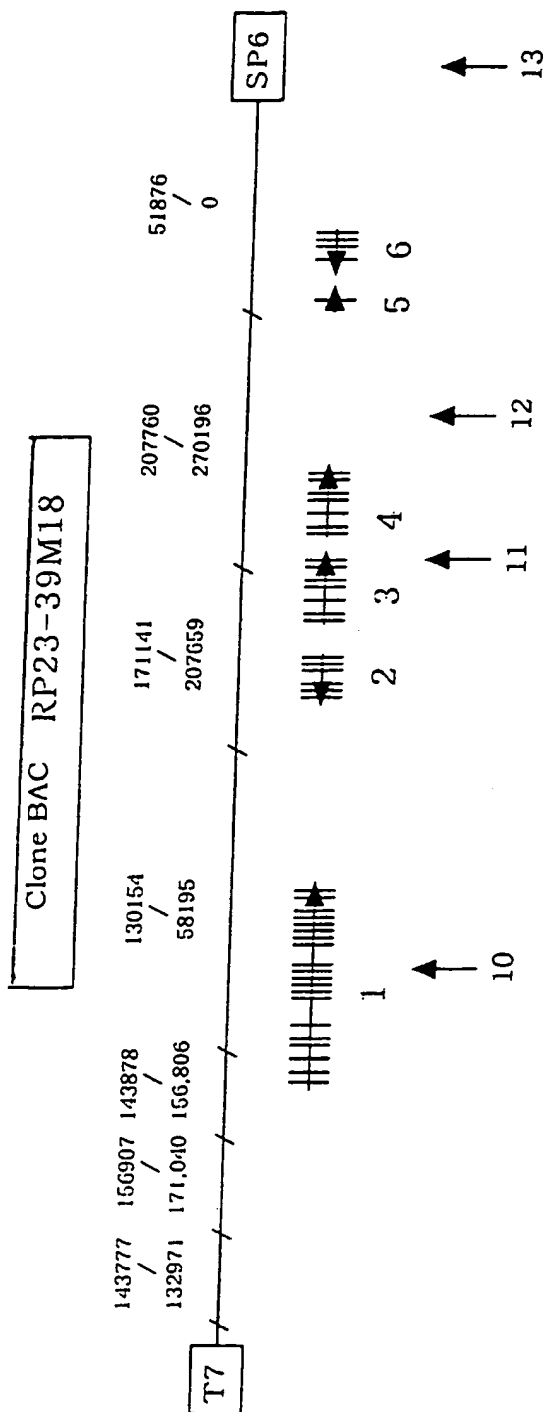
FIGURE 10

17/23

Phénotype <i>Flv</i> *	Survivants	Morts	Total
Résistant (<i>Flv^r/Flv^s</i>)	108	0	108 (55%)
Sensible (<i>Flv^s/Flv^s</i>)	21	74	95 (45%)
Total	129 (66%)	74 (34%)	203

* un allèle *Flv^r* est suffisant pour conférer la résistance

Figure 12



1) Raphiline 3A (P47708) 2) nouveau 2'5'OAS1 (AA794503+AA536864) 3) 2'5'OAS1.L3 (M33863) 4) 2'5'OAS1.L2 (X58077) 5) Retrovirus.. (P11365), GAG-EPMA (P03975) 6) 2'5'OAS1.L1 (M55982)	10) 698128; marqueur STS (Rab3A) 11) 302170; marqueur STS (OSA1) 12) D5Mit368; marqueurs SLP 13) 39M18SP6
---	--

FIGURE 13

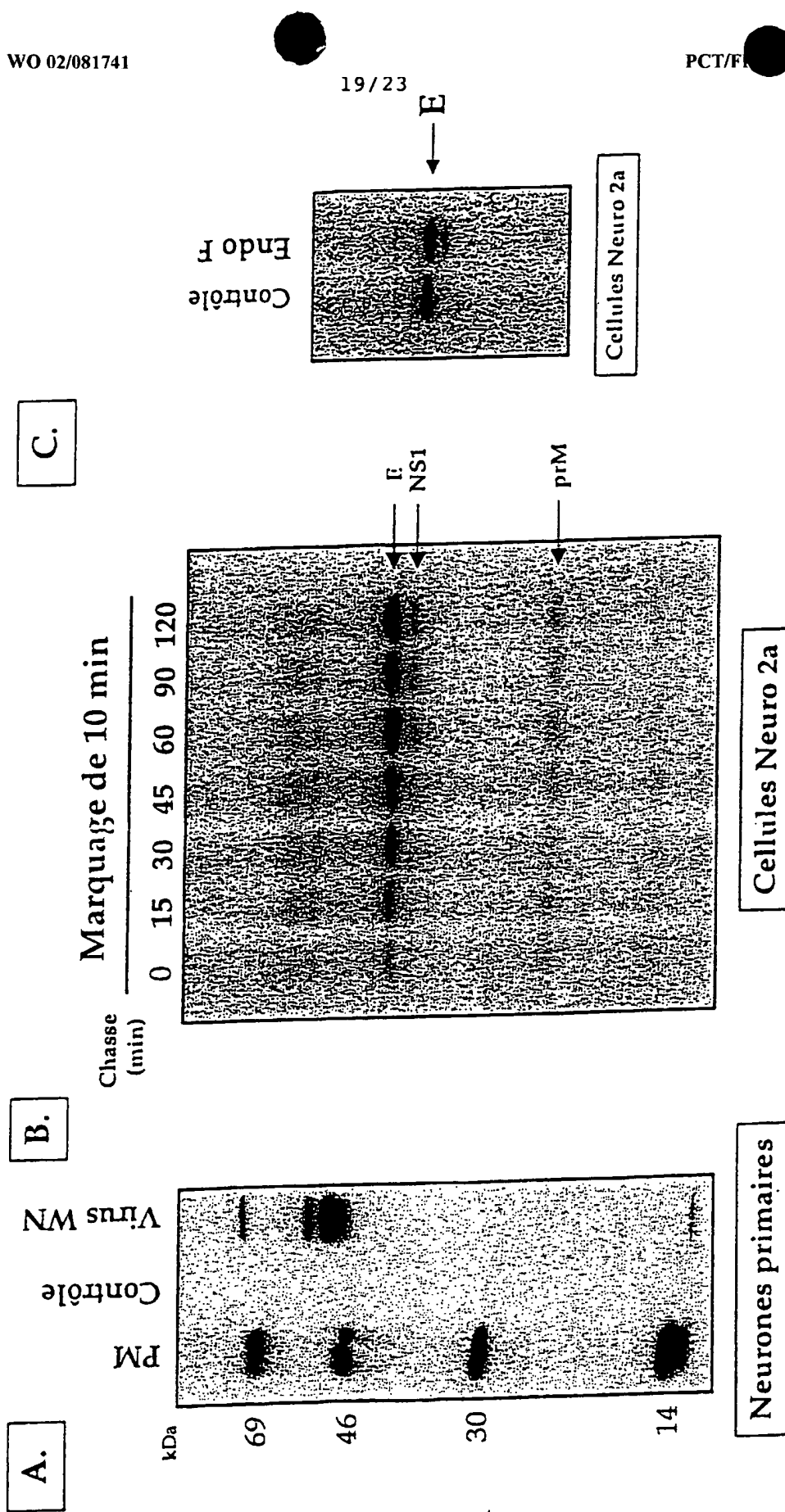
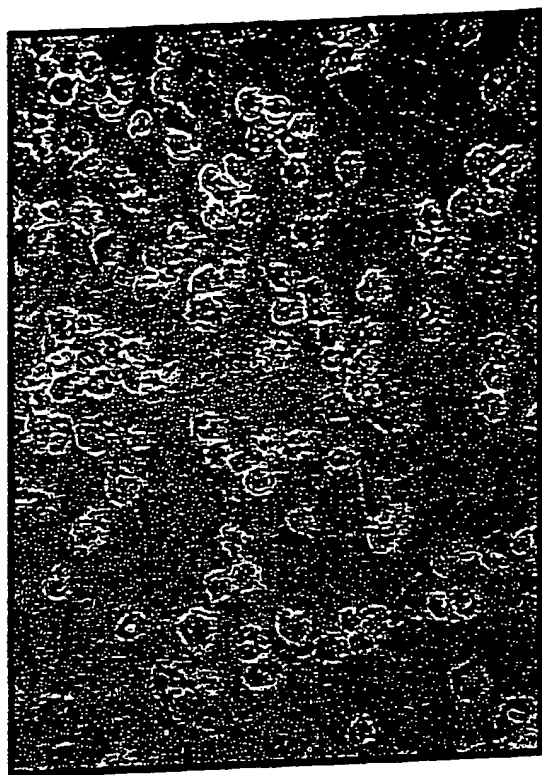
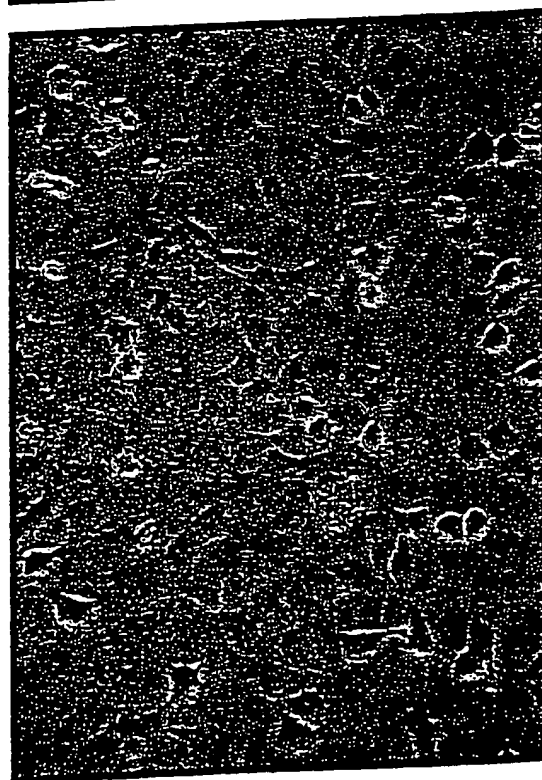


FIGURE 14



Souche IS-98-ST1 (m.i. de 4)
24 h d 'infection

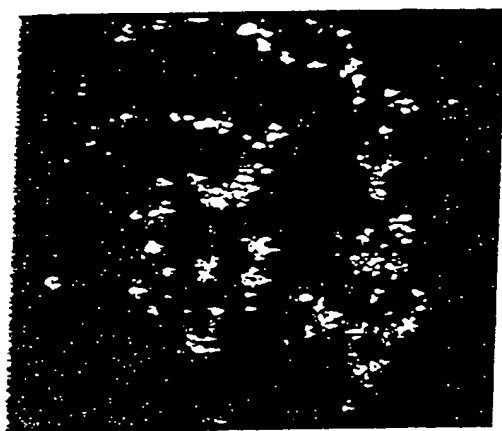


Contrôle

FIGURE 15

21/23

4 UFF/cell.

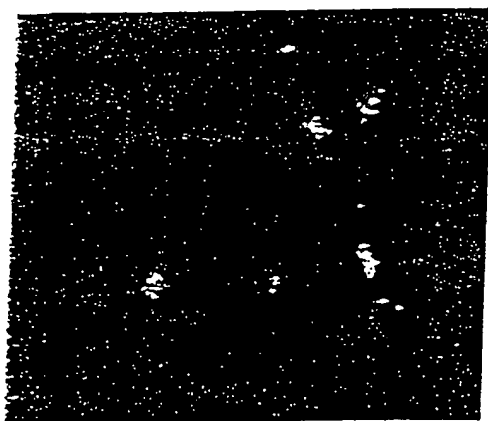


Contrôle



INF- α (20 UI/ml)

0,1 UFF/cell.



Contrôle



INF- α (10 UI/ml)

FIGURE 16

22/23

> Exon 1

AGGCTTGGATGGGGAGGTACCTGTTGAGAAAGCCCTAACGCCATTGGCTGCTCGGGCCTG
 GATGATTTGCATATCCGCGCCCTTCCCGGGAAATGGAACTGAAAGTCCCATTCTGCTT
 CAGCCAGCCTAGGAGACACAGGACCTGCTGGCTGCAGAGGTAAAAGCTGGACCTAGGA
 TGGAGCAGGATCTGAGGAGCATCCCGGCCTCGAAGCTTGATAAGTTCATAGAGAAC
 CATCTCCCGGACACCAGCTTCTGTGCTGACCTCAGAGAAGTCATAGATGCCCTGTG
 TGCTCTCCTGAAGGACAGATCCTTCCGGGGCCCCGTCCGCCGAATGAGGGCCTCT
 AAAGGGGTCAAGGTGAGCCTTCTCAGCCTGAGCTGGCCGAGATGAGGTGGGACAGG
 ACCTTTCAGAAGCCAGGCTGCAACCCTGATCCCTCCTCTTAATTCTGATCACAGCTGGCGA
 TGGGTTCTTCCCCCAAGTCCACATCTGTAATTGGAGAAGGAGCCTCAGCTACAGTTTAT
 GTTCCCCACTCCAGGCCATCTCCATTCAGAGTCGGGGAACTGAGGCCCAAGAATGGC

> Exon 2

TCAGGGCTGACTAATATTCTCTGCACTAGCAGCTTGCTCCATCCATCCAACCATCCA
 TCCATCCATCCATCCATCTATCACCTTGGTGATGGACTCCAGTTCAGCATCTGGACCTCCC
 AGCATGGCTCTCCCTGTCTTTGTCAATTCTTTCTTGTCTTTGAGGGCAAAGGCACCGC
 GCTCAAGGGCAGGTCAGACGCTGACCTGGTGGTGTTCCTTAACAATCTCACCAGCT
 TTGAGGATCAGTTAAACCAACAGGGAGTGTGATTAAGGAAATTAAGAAACAGCTG
 TGCGAGGTTGAGCATGAGAGACCTTGTGGAGTGAAGTTTGAGGTCCACAGTTTAA
 GGAGTCCCAACTCCCGGGCTCTGAGCTTCAAGCTGAGCGCCCCCGACCTGCTGAA
 GGAGGTGAAGTTTGATGTGCTGCCAGCCTATGATTACTGGGTAAGGCAGCCTGCCA
 GAGAGCCTCAGCTACCCCTTCTGCATGCCTTCATCCTCCCTTCTAGTTCCACCTCTGTGTG
 TGTGCACATGTGTGTGGTGGTGTGATCCGGAACCATGGC

> Exon 3

TGGGAGCAGGGACTTTGTAGTGACAGCATGTGTGCAGTCCTAAGGAGTGGTGGGGCAG
 ATGGGTATATGCGGAGCGATGCTGCTGGGGAAAAACGCTGCCTGTTGCTGAACCCCA
 TAGGCTGTGATTCTGGACCTGGGGAGACTGAGGAACCTCAGAGAGGGGCTGAGCTGTTT
 CACTGAATGAAATCTTGTTTACAGATCATCTTAACATCCTCAAGAAGCCTAACCAACA
 ATTCTACGCCAATCTCATCAGTGGGCGTACCCCGCCGGGGAAGGAGGGCAAGTTAT
 CGATCTGCTTTATGGGGCTTCAGAAAGTACTTCTGAAGTGTGCGCCCAACCAAGCTG
 AAGCGCCTCATCCGCTGGTCACGCACTGGTACCAACTGGTAAGGCATTGGCCTGGT
 CACTGTGCGCTTCATCTGTACAGAGTGTAGACACGGGGCAGAGAAAATGTGTACTCTAGT
 TCGGCGTCTATTGAGGTAATAACTCTATCACATGACCAAAACGCTCCTGGGAAGGAAGGA
 TATTCTTATGTTTACATCTAAGGCACACTCCATCATTTAGGGGAAGTCAGAGGAATAACT

FIGURE 17

> Exon 4 (terminal)

GCTGTCACTCATGGTGTCTGATTCTCCTGGGCCTGTTAGTTACGTGCTCAGTCTGTGG
F5 (SEQIDNO:19)
AGCACAAGTGTCCATCTCAACCACTCAGCTCTTGGCTATAGCACAAGAAGGCATAGACT
GAATTCAAGTGACTGGGTGTGACAGTGTGACAATAACCTGTCATGACATCATCAGGTG
GCAGCTGAATTGAGCCTGTGAGAAGCAGCCCCACACACAACCACCAGGAACCCCGTTTT
F6 (SEQIDNO:22)
TTTCTCATTGCTTTCTCTTCTCAGTGTAAAGGAGAAGCTGGGGGACCCGCTGCCCCCA
CAGTATGCCCTGGAGCTGCTCACAGTCTATGCCTGGGAGTATGGGAGTCGAGTAAC
(SEQIDNO:20) R7
TAAATTCAACACAGCCCAGGGCTTCTGAACCGTCTTGGAACTGGTCACCAAGTACAAA
(SEQIDNO:21)
CAGCTTCAAATCTACTGGACAGTGTATTATGACTTTCGACATCAAGAGGTCTCTGAATACC
R8
TGCACCAACAGCTCAAAAAA

FIGURE 17 (suite)

LISTE DE SEQUENCES

<110> INSTITUT PASTEUR
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)
GUENET, Jean-Louis
MASHIMO, Tomoji
SIMON-CHAZOTTES, Dominique
MONTAGUTELLI, Xavier
FRENKIEL, Marie-Pascale
DESPRES, Philippe
DEUBEL, Vincent
BONHOMME, François
LUCAS, Marianne

<120> Utilisation des produits des gènes de la famille 2'-5' oligoadénylate synthétase (OAS) pour le criblage d'antiviraux et la détection de la sensibilité à l'infection par les Flaviviridae.

<130> F226/CAS92 PCT

<140>

<141>

<160> 32

<170> PatentIn Ver. 2.1

<210> 1

<211> 11029

<212> ADN

<213> Flavivirus sp.

<220>

<221> CDS

<222> (97)..(10395)

<400> 1

```

agtagttcgc ctgtgtgagc tgacaaactt agtagtggtt gtagaggatta acaacaatta 60
acacagtgcg agctgtttct tagcacgaag atctcg atg tct aag aaa cca gga 114
                               Met Ser Lys Lys Pro Gly
                               1           5

ggg ccc ggc aag agc cgg gct gtc aat atg cta aaa cgc gga atg ccc 162
Gly Pro Gly Lys Ser Arg Ala Val Asn Met Leu Lys Arg Gly Met Pro
          10           15           20

cgc gtg ttg tcc ttg att gga ctg aag agg gct atg ttg agc ctg atc 210
Arg Val Leu Ser Leu Ile Gly Leu Lys Arg Ala Met Leu Ser Leu Ile
          25           30           35

gac ggc aag ggg cca ata cga ttt gtg ttg gct ctc ttg gcg ttc ttc 258
Asp Gly Lys Gly Pro Ile Arg Phe Val Leu Ala Leu Leu Ala Phe Phe
          40           45           50

agg ttc aca gca att gct ccg acc cga gca gtg ctg gat cga tgg aga 306
Arg Phe Thr Ala Ile Ala Pro Thr Arg Ala Val Leu Asp Arg Trp Arg
          55           60           65           70

ggt gtg aat aaa caa aca gcg atg aaa cac ctt ctg agt ttt aag aag 354
Gly Val Asn Lys Gln Thr Ala Met Lys His Leu Leu Ser Phe Lys Lys
          75           80           85

```

gaa cta ggg acc ttg acc agt gct atc aat cgg cgg agc tca aaa caa	402
Glu Leu Gly Thr Leu Thr Ser Ala Ile Asn Arg Arg Ser Ser Lys Gln	
90 95 100	
aag aaa aga gga gga aag acc gga att gca gtc atg att ggc ctg atc	450
Lys Lys Arg Gly Gly Lys Thr Gly Ile Ala Val Met Ile Gly Leu Ile	
105 110 115	
gcc agc gta gga gca gtt acc ctc tct aac ttc .caa ggg aag gtg atg	498
Ala Ser Val Gly Ala Val Thr Leu Ser Asn Phe Gln Gly Lys Val Met	
120 125 130	
atg acg gta aat gct act gac gtc aca gat gtc atc acg att cca aca	546
Met Thr Val Asn Ala Thr Asp Val Thr Asp Val Ile Thr Ile Pro Thr	
135 140 145 150	
gct gct gga aag aac cta tgc att gtc aga gca atg gat gtg gga tac	594
Ala Ala Gly Lys Asn Leu Cys Ile Val Arg Ala Met Asp Val Gly Tyr	
155 160 165	
atg tgc gat gat act atc act tat gaa tgc cca gtg ctg tgc gct ggt	642
Met Cys Asp Asp Thr Ile Thr Tyr Glu Cys Pro Val Leu Ser Ala Gly	
170 175 180	
aat gat cca gaa gac atc gac tgt tgg tgc aca aag tca gca gtc tac	690
Asn Asp Pro Glu Asp Ile Asp Cys Trp Cys Thr Lys Ser Ala Val Tyr	
185 190 195	
gtc agg tat gga aga tgc acc aag aca cgc cac tca aga cgc agt cgg	738
Val Arg Tyr Gly Arg Cys Thr Lys Thr Arg His Ser Arg Arg Ser Arg	
200 205 210	
agg tca ctg aca gtg cag aca cac gga gaa agc act cta gcg aac aag	786
Arg Ser Leu Thr Val Gln Thr His Gly Glu Ser Thr Leu Ala Asn Lys	
215 220 225 230	
aag ggg gct tgg atg gac agc acc aag gcc aca agg tat ttg gta aaa	834
Lys Gly Ala Trp Met Asp Ser Thr Lys Ala Thr Arg Tyr Leu Val Lys	
235 240 245	
aca gaa tca tgg atc ttg agg aac cct gga tat gcc ctg gtg gca gcc	882
Thr Glu Ser Trp Ile Leu Arg Asn Pro Gly Tyr Ala Leu Val Ala Ala	
250 255 260	
gtc att ggt tgg atg ctt ggg agc aac acc atg cag aga gtt gtg ttt	930
Val Ile Gly Trp Met Leu Gly Ser Asn Thr Met Gln Arg Val Val Phe	
265 270 275	
gtc gtg cta ttg ctt ttg gtg gcc cca gct tac agc ttt aac tgc ctt	978
Val Val Leu Leu Leu Val Ala Pro Ala Tyr Ser Phe Asn Cys Leu	
280 285 290	
gga atg agc aac aga gac ttc ttg gaa gga gtg tct gga gca aca tgg	1026
Gly Met Ser Asn Arg Asp Phe Leu Glu Gly Val Ser Gly Ala Thr Trp	
295 300 305 310	
gtg gat ttg gtt ctc gaa ggc gac agc tgc gtg act atc atg tct aag	1074
Val Asp Leu Val Leu Glu Gly Asp Ser Cys Val Thr Ile Met Ser Lys	
315 320 325	

gac aag cct acc atc gat gtg aag atg atg aat atg gag gcg gcc aac	1122
Asp Lys Pro Thr Ile Asp Val Lys Met Met Asn Met Glu Ala Ala Asn	
330 335 340	
ctg gca gag gtc cgc agt tat tgc tat ttg gct acc gtc agc gat ctc	1170
Leu Ala Glu Val Arg Ser Tyr Cys Tyr Leu Ala Thr Val Ser Asp Leu	
345 350 355	
tcc acc aaa gct gcg tgc ccg acc atg gga gaa gct cac aat gac aaa	1218
Ser Thr Lys Ala Ala Cys Pro Thr Met Gly Glu Ala His Asn Asp Lys	
360 365 370	
cgt gct gac cca gct ttt gtg tgc aga caa gga gtg gtg gac agg ggc	1266
Arg Ala Asp Pro Ala Phe Val Cys Arg Gln Gly Val Val Asp Arg Gly	
375 380 385 390	
tgg ggc aac ggc tgc gga cta ttt ggc aaa gga agc att gac aca tgc	1314
Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Ile Asp Thr Cys	
395 400 405	
gcc aaa ttt gcc tgc tct acc aag gca ata gga aga acc atc ttg aaa	1362
Ala Lys Phe Ala Cys Ser Thr Lys Ala Ile Gly Arg Thr Ile Leu Lys	
410 415 420	
gag aat atc aag tac gaa gtg gcc att ttt gtc cat gga cca act act	1410
Glu Asn Ile Lys Tyr Glu Val Ala Ile Phe Val His Gly Pro Thr Thr	
425 430 435	
gtg gag tcg cac gga aac tac tcc aca cag gtt gga gcc act cag gca	1458
Val Glu Ser His Gly Asn Tyr Ser Thr Gln Val Gly Ala Thr Gln Ala	
440 445 450	
ggg aga ttc agc atc act cct gcg gcg cct tca tac aca cta aag ctt	1506
Gly Arg Phe Ser Ile Thr Pro Ala Ala Pro Ser Tyr Thr Leu Lys Leu	
455 460 465 470	
gga gaa tat gga gag gtg aca gtg gac tgt gaa cca cgg tca ggg att	1554
Gly Glu Tyr Gly Glu Val Thr Val Asp Cys Glu Pro Arg Ser Gly Ile	
475 480 485	
gac acc aat gca tac tac gtg atg act gtt gga aca aag acg ttc ttg	1602
Asp Thr Asn Ala Tyr Tyr Val Met Thr Val Gly Thr Lys Thr Phe Leu	
490 495 500	
gtc cat cgt gag tgg ttc atg gac ctc aac ctc cct tgg agc agt gct	1650
Val His Arg Glu Trp Phe Met Asp Leu Asn Leu Pro Trp Ser Ser Ala	
505 510 515	
gga agt act gtg tgg agg aac aga gag acg tta atg gag ttt gag gaa	1698
Gly Ser Thr Val Trp Arg Asn Arg Glu Thr Leu Met Glu Phe Glu Glu	
520 525 530	
cca cac gcc acg aag cag tct gtg ata gca ttg ggc tca caa gag gga	1746
Pro His Ala Thr Lys Gln Ser Val Ile Ala Leu Gly Ser Gln Glu Gly	
535 540 545 550	
gct ctg cat caa gct ttg gct gga gcc att cct gtg gaa ttt tca agc	1794
Ala Leu His Gln Ala Leu Ala Gly Ala Ile Pro Val Glu Phe Ser Ser	
555 560 565	

aac act gtc aag ttg acg tcg ggt cat ttg aag tgt aga gtn aag atg Asn Thr Val Lys Leu Thr Ser Gly His Leu Lys Cys Arg Val Lys Met 570 575 580	1842
gaa aaa ttg cag ttg aag gga aca acc tat ggc gtc tgt tca aag gct Glu Lys Leu Gln Leu Lys Gly Thr Tyr Gly Val Cys Ser Lys Ala 585 590 595	1890
ttc aag ttt ctt ggg act ccc gca gac aca ggt cac ggc act gtg gtg Phe Lys Phe Leu Gly Thr Pro Ala Asp Thr Gly His Gly Thr Val Val 600 605 610	1938
ttg gaa ttg cag tac act ggc acg gat gga cct tgc aaa gtt cct atc Leu Glu Leu Gln Tyr Thr Gly Thr Asp Gly Pro Cys Lys Val Pro Ile 615 620 625 630	1986
tcg tca gtg gct tca ttg aac gac cta acg cca gtg ggc aga ttg gtc Ser Ser Val Ala Ser Leu Asn Asp Leu Thr Pro Val Gly Arg Leu Val 635 640 645	2034
act gtc aac cct ttt gtt tca gtg gcc acg gcc aac gct aag gtc ctg Thr Val Asn Pro Phe Val Ser Val Ala Thr Ala Asn Ala Lys Val Leu 650 655 660	2082
att gaa ttg gaa cca ccc ttt gga gac tca tac ata gtg gtg ggc aga Ile Glu Leu Glu Pro Pro Phe Gly Asp Ser Tyr Ile Val Val Gly Arg 665 670 675	2130
gga gaa caa cag att aat cac cat tgg cac aag tct gga agc agc att Gly Glu Gln Gln Ile Asn His His Trp His Lys Ser Gly Ser Ser Ile 680 685 690	2178
ggc aaa gcc ttt aca acc acc ctc aaa gga gcg cag aga cta gcc gct Gly Lys Ala Phe Thr Thr Thr Leu Lys Gly Ala Gln Arg Leu Ala Ala 695 700 705 710	2226
cta gga gac aca gct tgg gac ttt gga tca gtt gga ggg gtg ttc acc Leu Gly Asp Thr Ala Trp Asp Phe Gly Ser Val Gly Gly Val Phe Thr 715 720 725	2274
tca gtt ggg aag gct gtc cat caa gtg ttc gga gga gca ttc cgc tca Ser Val Gly Lys Ala Val His Gln Val Phe Gly Gly Ala Phe Arg Ser 730 735 740	2322
ctg ttc gga ggc atg tcc tgg ata acg caa gga ttg ctg ggg gct ctc Leu Phe Gly Gly Met Ser Trp Ile Thr Gln Gly Leu Leu Gly Ala Leu 745 750 755	2370
ctg ttg tgg atg ggc atc aat gct cgt gat agg tcc ata gct ctc acg Leu Leu Trp Met Gly Ile Asn Ala Arg Asp Arg Ser Ile Ala Leu Thr 760 765 770	2418
ttt ctc gca gtt gga gga gtt ctg ctc ttc ctc tcc gtg aac gtg cac Phe Leu Ala Val Gly Gly Val Leu Leu Phe Leu Ser Val Asn Val His 775 780 785 790	2466
gct gac act ggg tgt gcc ata gac atc agc cgg caa gag ctg aga tgt Ala Asp Thr Gly Cys Ala Ile Asp Ile Ser Arg Gln Glu Leu Arg Cys 795 800 805	2514

gga aat gga gtg ttc ata cac aat gat gtg gag gct tgg atg gac cgg	2562
Gly Asn Gly Val Phe Ile His Asn Asp Val Glu Ala Trp Met Asp Arg	
810 815 820	
tac aag tat tac cct gaa acg cca caa ggc cta gcc aag atc att cag	2610
Tyr Lys Tyr Tyr Pro Glu Thr Pro Gln Gly Leu Ala Lys Ile Ile Gln	
825 830 835	
aaa gct cat aag gaa gga gtg tgc ggt cta cga tca gtt tcc aga ctg	2658
Lys Ala His Lys Glu Gly Val Cys Gly Leu Arg Ser Val Ser Arg Leu	
840 845 850	
gag cat caa atg tgg gaa gca gtg aag gac gag ctg aac act ctt ttg	2706
Glu His Gln Met Trp Glu Ala Val Lys Asp Glu Leu Asn Thr Leu Leu	
855 860 865 870	
aag gag aat ggt gtg gac ctt agt gtc gtg gtt gag aaa cag gag gga	2754
Lys Glu Asn Gly Val Asp Leu Ser Val Val Val Glu Lys Gln Glu Gly	
875 880 885	
atg tac aag tca gca cct aaa cgc ctc acc gcc acc acg gaa aaa ttg	2802
Met Tyr Lys Ser Ala Pro Lys Arg Leu Thr Ala Thr Thr Glu Lys Leu	
890 895 900	
gaa att ggc tgg aag gcc tgg gga aag agt att tta ttt gca cca gaa	2850
Glu Ile Gly Trp Lys Ala Trp Gly Lys Ser Ile Leu Phe Ala Pro Glu	
905 910 915	
ctc gcc aac aac acc ttt gtg gtt gat ggt ccg gag acc aag gaa tgt	2898
Leu Ala Asn Asn Thr Phe Val Val Asp Gly Pro Glu Thr Lys Glu Cys	
920 925 930	
ccg act cag aat cgc gct tgg aat agc tta gaa gtg gag gat ttt gga	2946
Pro Thr Gln Asn Arg Ala Trp Asn Ser Leu Glu Val Glu Asp Phe Gly	
935 940 945 950	
ttt ggt ctc acc agc act cgg atg ttc ctg aag gtc aga gag agc aac	2994
Phe Gly Leu Thr Ser Thr Arg Met Phe Leu Lys Val Arg Glu Ser Asn	
955 960 965	
aca act gaa tgt gac tcg aag atc att gga acg gct gtc aag aac aac	3042
Thr Thr Glu Cys Asp Ser Lys Ile Ile Gly Thr Ala Val Lys Asn Asn	
970 975 980	
ttg gcg atc cac agt gac ctg tcc tat tgg att gaa agc agg ctc aat	3090
Leu Ala Ile His Ser Asp Leu Ser Tyr Trp Ile Glu Ser Arg Leu Asn	
985 990 995	
gat acg tgg aag ctt gaa agg gca gtt ctg ggt gaa gtc aaa tca tgt	3138
Asp Thr Trp Lys Leu Glu Arg Ala Val Leu Gly Glu Val Lys Ser Cys	
1000 1005 1010	
acg tgg cct gag acg cat acc ttg tgg ggc gat gga atc ctt gag agt	3186
Thr Trp Pro Glu Thr His Thr Leu Trp Gly Asp Gly Ile Leu Glu Ser	
1015 1020 1025 1030	
gac ttg ata ata cca gtc aca ctg gcg gga cca cga agc aat cac aat	3234
Asp Leu Ile Ile Pro Val Thr Leu Ala Gly Pro Arg Ser Asn His Asn	
1035 1040 1045	

cgg aga cct ggg tac aag aca caa aac cag ggc cca tgg gac gaa ggc 3282
 Arg Arg Pro Gly Tyr Lys Thr Gln Asn Gln Gly Pro Trp Asp Glu Gly
 1050 1055 1060

cgg gta gag att gac ttc gat tac tgc cca gga act acg gtc acc ctg 3330
 Arg Val Glu Ile Asp Phe Asp Tyr Cys Pro Gly Thr Thr Val Thr Leu
 1065 1070 1075

agt gag agc tgc gga cac cgt gga cct gcc act cgc acc acc aca gag 3378
 Ser Glu Ser Cys Gly His Arg Gly Pro Ala Thr Arg Thr Thr Thr Glu
 1080 1085 1090

agc gga aag ttg ata aca gat tgg tgc tgc agg agc tgc acc tta cca 3426
 Ser Gly Lys Leu Ile Thr Asp Trp Cys Cys Arg Ser Cys Thr Leu Pro
 1095 1100 1105 1110

cca ctg cgc tac caa act gac agc ggc tgt tgg tat ggt atg gag atc 3474
 Pro Leu Arg Tyr Gln Thr Asp Ser Gly Cys Trp Tyr Gly Met Glu Ile
 1115 1120 1125

aga cca cag aga cat gat gaa aag acc ctc gtg cag tca caa gtg aat 3522
 Arg Pro Gln Arg His Asp Glu Lys Thr Leu Val Gln Ser Gln Val Asn
 1130 1135 1140

gct tat aat gct gat atg att gac cct ttt cag ttg ggc ctt ctg gtc 3570
 Ala Tyr Asn Ala Asp Met Ile Asp Pro Phe Gln Leu Gly Leu Leu Val
 1145 1150 1155

gtg ttc ttg gcc acc cag gag gtc ctt cgc aag agg tgg aca gcc aag 3618
 Val Phe Leu Ala Thr Gln Glu Val Leu Arg Lys Arg Trp Thr Ala Lys
 1160 1165 1170

atc agc atg cca gct ata ctg att gct ctg cta gtc ctg gtg ttt ggg 3666
 Ile Ser Met Pro Ala Ile Leu Ile Ala Leu Leu Val Leu Val Phe Gly
 1175 1180 1185 1190

ggc att act tac act gat gtg tta cgc tat gtc atc ttg gtg ggg gca 3714
 Gly Ile Thr Tyr Thr Asp Val Leu Arg Tyr Val Ile Leu Val Gly Ala
 1195 1200 1205

gct ttc gca gaa tct aat tcg gga gga gac gtg gta cac ttg gcg ctc 3762
 Ala Phe Ala Glu Ser Asn Ser Gly Gly Asp Val Val His Leu Ala Leu
 1210 1215 1220

atg gcg acc ttc aag ata caa cca gtg ttt atg gtg gca tcg ttt ctc 3810
 Met Ala Thr Phe Lys Ile Gln Pro Val Phe Met Val Ala Ser Phe Leu
 1225 1230 1235

aaa gcg aga tgg acc aac cag gag aac att ttg ttg atg ttg gcg gct 3858
 Lys Ala Arg Trp Thr Asn Gln Glu Asn Ile Leu Leu Met Leu Ala Ala
 1240 1245 1250

gtt ttc ttt caa atg gct tat cac gat gcc cgc caa att ctg ctc tgg 3906
 Val Phe Phe Gln Met Ala Tyr His Asp Ala Arg Gln Ile Leu Leu Trp
 1255 1260 1265 1270

gag atc cct gat gtg ttg aat tca ctg gcg gta gct tgg atg ata ctg 3954
 Glu Ile Pro Asp Val Leu Asn Ser Leu Ala Val Ala Trp Met Ile Leu
 1275 1280 1285

aga gcc ata aca ttc aca acg aca tca aat gtg gtc gtc ccg ctg cta Arg Ala Ile Thr Phe Thr Thr Thr Ser Asn Val Val Val Pro Leu Leu 1290 1295 1300	4002
gcc ctg cta aca ccc cgg ctg aga tgc ttg aat ctg gat gtg tac agg Ala Leu Leu Thr Pro Arg Leu Arg Cys Leu Asn Leu Asp Val Tyr Arg 1305 1310 1315	4050
ata ctg ctg ttg atg gtc gga ata ggc agc ttg atc agg gag aag agg Ile Leu Leu Leu Met Val Gly Ile Gly Ser Leu Ile Arg Glu Lys Arg 1320 1325 1330	4098
agt gca gct gca aaa aag aaa gga gca agt ctg cta tgc ttg gct cta Ser Ala Ala Ala Lys Lys Lys Gly Ala Ser Leu Leu Cys Leu Ala Leu 1335 1340 1345 1350	4146
gcc tca aca gga ctt ttc aac ccc atg atc ctt gct gct gga ctg att Ala Ser Thr Gly Leu Phe Asn Pro Met Ile Leu Ala Ala Gly Leu Ile 1355 1360 1365	4194
gca tgt gat ccc aac cgt aaa cgc gga tgg ccc gca act gaa gtg atg Ala Cys Asp Pro Asn Arg Lys Arg Gly Trp Pro Ala Thr Glu Val Met 1370 1375 1380	4242
aca gct gtc ggc cta atg ttt gcc atc gtc gga ggg ctg gca gag ctt Thr Ala Val Gly Leu Met Phe Ala Ile Val Gly Gly Leu Ala Glu Leu 1385 1390 1395	4290
gac att gac tcc atg gcc att cca atg act atc gcg ggg ctc atg ttt Asp Ile Asp Ser Met Ala Ile Pro Met Thr Ile Ala Gly Leu Met Phe 1400 1405 1410	4338
gct gct ttc gtg att tct ggg aaa tca aca gat atg tgg att gag aga Ala Ala Phe Val Ile Ser Gly Lys Ser Thr Asp Met Trp Ile Glu Arg 1415 1420 1425 1430	4386
acg gcg gac att tcc tgg gaa agt gat gca gaa att aca ggc tcg agc Thr Ala Asp Ile Ser Trp Glu Ser Asp Ala Glu Ile Thr Gly Ser Ser 1435 1440 1445	4434
gaa aga gtt gat gtt cgg ctt gat gat ggt gaa aac ttc cag ctc atg Glu Arg Val Asp Val Arg Leu Asp Asp Gly Glu Asn Phe Gln Leu Met 1450 1455 1460	4482
aat gat cca gga gca cct tgg aag ata tgg atg ctc aga atg gtc tgt Asn Asp Pro Gly Ala Pro Trp Lys Ile Trp Met Leu Arg Met Val Cys 1465 1470 1475	4530
ctc gcg att agt gcg tac acc ccc tgg gca atc ttg ccc tca gta gtt Leu Ala Ile Ser Ala Tyr Thr Pro Trp Ala Ile Leu Pro Ser Val Val 1480 1485 1490	4578
gga ttt tgg ata act ctc caa tac aca aag aga gga ggt gtg ttg tgg Gly Phe Trp Ile Thr Leu Gln Tyr Thr Lys Arg Gly Gly Val Leu Trp 1495 1500 1505 1510	4626
gac act ccc tca cca aag gag tac aaa aag ggg gac acg acc acc ggc Asp Thr Pro Ser Pro Lys Glu Tyr Lys Lys Gly Asp Thr Thr Thr Gly 1515 1520 1525	4674

gtc tac agg atc atg act cgt ggg ctg ctc ggc agt tat caa gca gga 4722
 Val Tyr Arg Ile Met Thr Arg Gly Leu Leu Gly Ser Tyr Gln Ala Gly
 1530 1535 1540

gcg ggc gtg atg gtt gaa ggt gtt ttc cac acc ctt tgg cat aca aca 4770
 Ala Gly Val Met Val Glu Gly Val Phe His Thr Leu Trp His Thr Thr
 1545 1550 1555

aaa gga gcc gct ttg atg agc gga gag ggc cgc ctg gac cca tac tgg 4818
 Lys Gly Ala Ala Leu Met Ser Gly Glu Gly Arg Leu Asp Pro Tyr Trp
 1560 1565 1570

ggc agt gtc aag gag gat cga ctt tgt tac gga gga ccc tgg aaa ttg 4866
 Gly Ser Val Lys Glu Asp Arg Leu Cys Tyr Gly Gly Pro Trp Lys Leu
 1575 1580 1585 1590

cag cac aag tgg aac ggg cag gat gag gtg cag atg att gtg gtg gaa 4914
 Gln His Lys Trp Asn Gly Gln Asp Glu Val Gln Met Ile Val Val Glu
 1595 1600 1605

cct ggc aag aac gtt aag aac gtc cag acg aaa cca ggg gtg ttc aaa 4962
 Pro Gly Lys Asn Val Lys Asn Val Gln Thr Lys Pro Gly Val Phe Lys
 1610 1615 1620

aca cct gaa gga gaa atc ggg gcc gtg act ttg gac ttc ccc act gga 5010
 Thr Pro Glu Gly Glu Ile Gly Ala Val Thr Leu Asp Phe Pro Thr Gly
 1625 1630 1635

aca tca ggc tca cca ata gtg gac aaa aac ggt gat gtg att ggg ctt 5058
 Thr Ser Gly Ser Pro Ile Val Asp Lys Asn Gly Asp Val Ile Gly Leu
 1640 1645 1650

tat ggc aat gga gtc ata atg ccc aac ggc tca tac ata agc gcg ata 5106
 Tyr Gly Asn Gly Val Ile Met Pro Asn Gly Ser Tyr Ile Ser Ala Ile
 1655 1660 1665 1670

gtg cag ggt gaa agg atg gat gag cca atc cca gcc gga ttc gaa cct 5154
 Val Gln Gly Glu Arg Met Asp Glu Pro Ile Pro Ala Gly Phe Glu Pro
 1675 1680 1685

gag atg ctg agg aaa aaa cag atc act gta ctg gat ctc cat ccc ggc 5202
 Glu Met Leu Arg Lys Lys Gln Ile Thr Val Leu Asp Leu His Pro Gly
 1690 1695 1700

gcc ggt aaa aca agg agg att ctg cca cag atc atc aaa gag gcc ata 5250
 Ala Gly Lys Thr Arg Arg Ile Leu Pro Gln Ile Ile Lys Glu Ala Ile
 1705 1710 1715

aac aga aga ctg aga aca gcc gtg cta gca cca acc agg gtt gtg gct 5298
 Asn Arg Arg Leu Arg Thr Ala Val Leu Ala Pro Thr Arg Val Val Ala
 1720 1725 1730

gct gag atg gct gaa gca ctg aga gga ctg ccc atc cgg tac cag aca 5346
 Ala Glu Met Ala Glu Ala Leu Arg Gly Leu Pro Ile Arg Tyr Gln Thr
 1735 1740 1745 1750

tcc gca gtg ccc aga gaa cat aat gga aat gag att gtt gat gtc atg 5394
 Ser Ala Val Pro Arg Glu His Asn Gly Asn Glu Ile Val Asp Val Met
 1755 1760 1765

tgt cat gct acc ctc acc cac agg ctg atg tct cct cac agg gtg ccg Cys His Ala Thr Leu Thr His Arg Leu Met Ser Pro His Arg Val Pro 1770 1775 1780	5442
aac tac aac ctg ttc gtg atg gat gag gct cat ttc acc gac cca gct Asn Tyr Asn Leu Phe Val Met Asp Glu Ala His Phe Thr Asp Pro Ala 1785 1790 1795	5490
agt atc gca gca aga ggt tac att tcc aca aag gtc gag cta ggg gag Ser Ile Ala Ala Arg Gly Tyr Ile Ser Thr Lys Val Glu Leu Gly Glu 1800 1805 1810	5538
gcg gcg gca ata ttc atg aca gcc acc cca cca ggc act tca gat cca Ala Ala Ala Ile Phe Met Thr Ala Thr Pro Pro Gly Thr Ser Asp Pro 1815 1820 1825 1830	5586
ttc cca gag tcc aat tca cca att tcc gac tta cag act gag atc ccg Phe Pro Glu Ser Asn Ser Pro Ile Ser Asp Leu Gln Thr Glu Ile Pro 1835 1840 1845	5634
gat cga gct tgg aac tct gga tac gaa tgg atc aca gaa tac acc ggg Asp Arg Ala Trp Asn Ser Gly Tyr Glu Trp Ile Thr Glu Tyr Thr Gly 1850 1855 1860	5682
aag acg gtt tgg ttt gtg cct agt gtc aag atg ggg aat gag att gcc Lys Thr Val Trp Phe Val Pro Ser Val Lys Met Gly Asn Glu Ile Ala 1865 1870 1875	5730
ctt tgc cta caa cgt gct gga aag aaa gta gtc caa ttg aac aga aag Leu Cys Leu Gln Arg Ala Gly Lys Lys Val Val Gln Leu Asn Arg Lys 1880 1885 1890	5778
tcg tac gag acg gag tac cca aaa tgt aag aac gat gat tgg gac ttt Ser Tyr Glu Thr Glu Tyr Pro Lys Cys Lys Asn Asp Asp Trp Asp Phe 1895 1900 1905 1910	5826
gtt atc aca aca gac ata tct gaa atg ggg gct aac ttc aag gcg agc Val Ile Thr Thr Asp Ile Ser Glu Met Gly Ala Asn Phe Lys Ala Ser 1915 1920 1925	5874
agg gtg att gac agc cgg aag agt gtg aaa cca acc atc ata aca gaa Arg Val Ile Asp Ser Arg Lys Ser Val Lys Pro Thr Ile Ile Thr Glu 1930 1935 1940	5922
gga gaa ggg aga gtg atc ctg gga gaa cca tct gca gtg aca gca gct Gly Glu Gly Arg Val Ile Leu Gly Glu Pro Ser Ala Val Thr Ala Ala 1945 1950 1955	5970
agt gcc gcc cag aga cgt gga cgt atc ggt aga aat ccg tcg caa gtt Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn Pro Ser Gln Val 1960 1965 1970	6018
ggt gat gag tac tgt tat ggg ggg cac acg aat gaa gac gac tcg aac Gly Asp Glu Tyr Cys Tyr Gly Gly His Thr Asn Glu Asp Asp Ser Asn 1975 1980 1985 1990	6066
ttc gcc cat tgg act gag gca cga atc atg ccg gac aac atc aac atg Phe Ala His Trp Thr Glu Ala Arg Ile Met Pro Asp Asn Ile Asn Met 1995 2000 2005	6114

cca aac gga ctg atc gct caa ttc tac caa cca gag cgt gag aag gta Pro Asn Gly Leu Ile Ala Gln Phe Tyr Gln Pro Glu Arg Glu Lys Val 2010 2015 2020	6162
tat acc atg gag ggg gaa tac cgg ctc aga gga gaa gag agg aaa aac Tyr Thr Met Glu Gly Glu Tyr Arg Leu Arg Gly Glu Glu Arg Lys Asn 2025 2030 2035	6210
ttt ctg gaa ctg ttg agg act gca gat ctg cca gtt tgg ctg gct tac Phe Leu Glu Leu Leu Arg Thr Ala Asp Leu Pro Val Trp Leu Ala Tyr 2040 2045 2050	6258
aag gtt gca gcg gct gga gtg tca tac cac gac cgg agg tgg tgc ttt Lys Val Ala Ala Ala Gly Val Ser Tyr His Asp Arg Arg Trp Cys Phe 2055 2060 2065 2070	6306
gat ggt cct agg aca aac aca att tta gaa gac aac aac gaa gtg gaa Asp Gly Pro Arg Thr Asn Thr Ile Leu Glu Asp Asn Asn Glu Val Glu 2075 2080 2085	6354
gtc atc acg aag ctt ggt gaa agg aag att ctg agg ccg cgc tgg att Val Ile Thr Lys Leu Gly Glu Arg Lys Ile Leu Arg Pro Arg Trp Ile 2090 2095 2100	6402
gac gcc agg gtg tac tcg gat cac cag gca cta aag gcg ttc aag gac Asp Ala Arg Val Tyr Ser Asp His Gln Ala Leu Lys Ala Phe Lys Asp 2105 2110 2115	6450
ttc gcc tcg gga aaa cgt tct cag ata ggg ctc att gag gtt ctg gga Phe Ala Ser Gly Lys Arg Ser Gln Ile Gly Leu Ile Glu Val Leu Gly 2120 2125 2130	6498
aag atg cct gag cac ttc atg ggg aag aca tgg gaa gca ctt gac acc Lys Met Pro Glu His Phe Met Gly Lys Thr Trp Glu Ala Leu Asp Thr 2135 2140 2145 2150	6546
atg tac gtt gtg gcc act gca gag aaa gga gga aga gct cac aga atg Met Tyr Val Val Ala Thr Ala Glu Lys Gly Gly Arg Ala His Arg Met 2155 2160 2165	6594
gcc ctg gag gaa ctg cca gat gct ctt cag aca att gcc ttg att gcc Ala Leu Glu Glu Leu Pro Asp Ala Leu Gln Thr Ile Ala Leu Ile Ala 2170 2175 2180	6642
tta ttg agt gtg atg acc atg gga gta ttc ttc ctc ctc atg cag cgg Leu Leu Ser Val Met Thr Met Gly Val Phe Phe Leu Leu Met Gln Arg 2185 2190 2195	6690
aag ggc att gga aag ata ggt ttg gga ggc gct gtc ttg gga gtc gcg Lys Gly Ile Gly Lys Ile Gly Leu Gly Gly Ala Val Leu Gly Val Ala 2200 2205 2210	6738
acc ttt ttc tgt tgg atg gct gaa gtt cca gga acg aag atc gcc gga Thr Phe Phe Cys Trp Met Ala Glu Val Pro Gly Thr Lys Ile Ala Gly 2215 2220 2225 2230	6786
atg ttg ctg ctc tcc ctt ctc ttg atg att gtg cta att cct gag cca Met Leu Leu Leu Ser Leu Leu Leu Met Ile Val Leu Ile Pro Glu Pro 2235 2240 2245	6834

gag aag caa cgt tcg cag aca gac aac cag cta gcc gtg ttc ctg att Glu Lys Gln Arg Ser Gln Thr Asp Asn Gln Leu Ala Val Phe Leu Ile 2250 2255 2260	6882
tgt gtc atg acc ctt gtg agc gca gtg gca gcc aac gag atg ggt tgg Cys Val Met Thr Leu Val Ser Ala Val Ala Ala Asn Glu Met Gly Trp 2265 2270 2275	6930
cta gac aag acc aag agt gac ata agc agt ttg ttt ggg caa aga att Leu Asp Lys Thr Lys Ser Asp Ile Ser Ser Leu Phe Gly Gln Arg Ile 2280 2285 2290	6978
gag gtc aag gag aat ttc agc atg gga gag ttt ctt ctg gac ttg agg Glu Val Lys Glu Asn Phe Ser Met Gly Glu Phe Leu Leu Asp Leu Arg 2295 2300 2305 2310	7026
ccg gca aca gcc tgg tca ctg tac gct gtg aca aca gcg gtc ctc act Pro Ala Thr Ala Trp Ser Leu Tyr Ala Val Thr Thr Ala Val Leu Thr 2315 2320 2325	7074
cca ctg cta aag cat ttg atc acg tca gat tac atc aac acc tca ttg Pro Leu Leu Lys His Leu Ile Thr Ser Asp Tyr Ile Asn Thr Ser Leu 2330 2335 2340	7122
acc tca ata aac gtt cag gca agt gca cta ttc aca ctc gcg cga ggc Thr Ser Ile Asn Val Gln Ala Ser Ala Leu Phe Thr Leu Ala Arg Gly 2345 2350 2355	7170
ttc ccc ttc gtc gat gtt gga gtg tcg gct ctc ctg cta gca gcc gga Phe Pro Phe Val Asp Val Gly Val Ser Ala Leu Leu Leu Ala Ala Gly 2360 2365 2370	7218
tgc tgg gga caa gtc acc ctc acc gtt acg gta aca gcg gca aca ctc Cys Trp Gly Gln Val Thr Leu Thr Val Thr Val Thr Ala Ala Thr Leu 2375 2380 2385 2390	7266
ctt ttt tgc cac tat gcc tac atg gtt ccc ggt tgg caa gct gag gca Leu Phe Cys His Tyr Ala Tyr Met Val Pro Gly Trp Gln Ala Glu Ala 2395 2400 2405	7314
atg cgc tca gcc cag cgg cgg aca gcg gcc gga atc atg aaa aac gct Met Arg Ser Ala Gln Arg Arg Thr Ala Ala Gly Ile Met Lys Asn Ala 2410 2415 2420	7362
gta gtg gat ggc atc gtg gcc acg gac gtc cca gaa tta gag cgc acc Val Val Asp Gly Ile Val Ala Thr Asp Val Pro Glu Leu Glu Arg Thr 2425 2430 2435	7410
aca ccc atc atg cag aag aaa gtt gga cag atc atg ctg atc ttg gtg Thr Pro Ile Met Gln Lys Lys Val Gly Gln Ile Met Leu Ile Leu Val 2440 2445 2450	7458
tct cta gct gca gta gta gtg aac ccg tct gtg aag aca gta cga gaa Ser Leu Ala Ala Val Val Val Asn Pro Ser Val Lys Thr Val Arg Glu 2455 2460 2465 2470	7506
gcc gga att ttg atc acg gcc gca gcg gtg acg ctt tgg gag aat gga Ala Gly Ile Leu Ile Thr Ala Ala Ala Val Thr Leu Trp Glu Asn Gly 2475 2480 2485	7554

gca agc tct gtt tgg aac gca aca act gcc atc gga ctc tgc cac atc Ala Ser Ser Val Trp Asn Ala Thr Thr Ala Ile Gly Leu Cys His Ile 2490 2495 2500	7602
atg cgt ggg ggt tgg ttg tca tgt cta tcc ata aca tgg aca ctc ata Met Arg Gly Gly Trp Leu Ser Cys Leu Ser Ile Thr Trp Thr Leu Ile 2505 2510 2515	7650
aag aac atg gaa aaa cca gga cta aaa aga ggt ggg gca aaa gga cgc Lys Asn Met Glu Lys Pro Gly Leu Lys Arg Gly Gly Ala Lys Gly Arg 2520 2525 2530	7698
acc ttg gga gag gtt tgg aaa gaa aga ctc aac cag atg aca aaa gaa Thr Leu Gly Glu Val Trp Lys Glu Arg Leu Asn Gln Met Thr Lys Glu 2535 2540 2545 2550	7746
gag ttc act agg tac cgc aaa gag gcc atc atc gaa gtc gat cgc tca Glu Phe Thr Arg Tyr Arg Lys Glu Ala Ile Ile Glu Val Asp Arg Ser 2555 2560 2565	7794
gcg gca aaa cac gcc agg aaa gaa ggc aat gtc act gga ggg cat tca Ala Ala Lys His Ala Arg Lys Glu Gly Asn Val Thr Gly Gly His Ser 2570 2575 2580	7842
gtc tct agg ggc aca gca aaa ctg aga tgg ctg gtc gaa cgg agg ttt Val Ser Arg Gly Thr Ala Lys Leu Arg Trp Leu Val Glu Arg Arg Phe 2585 2590 2595	7890
ctc gaa ccg gtc gga aaa gtg att gac ctt gga tgt gga aga ggc ggt Leu Glu Pro Val Gly Lys Val Ile Asp Leu Gly Cys Gly Arg Gly Gly 2600 2605 2610	7938
tgg tgt tac tat atg gca acc caa aaa aga gtc caa gaa gtc aga ggg Trp Cys Tyr Tyr Met Ala Thr Gln Lys Arg Val Gln Glu Val Arg Gly 2615 2620 2625 2630	7986
tac aca aag ggc ggt ccc gga cat gaa gag ccc caa cta gtg caa agt Tyr Thr Lys Gly Gly Pro Gly His Glu Glu Pro Gln Leu Val Gln Ser 2635 2640 2645	8034
tat gga tgg aac att gtc acc atg aag agt gga gtg gat gtg ttc tac Tyr Gly Trp Asn Ile Val Thr Met Lys Ser Gly Val Asp Val Phe Tyr 2650 2655 2660	8082
aga cct tct gag tgt tgt gac acc ctc ctt tgt gac atc gga gag tcc Arg Pro Ser Glu Cys Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser 2665 2670 2675	8130
tcg tca agt gct gag gtt gaa gag cat agg acg att cgt gtc ctt gaa Ser Ser Ser Ala Glu Val Glu Glu His Arg Thr Ile Arg Val Leu Glu 2680 2685 2690	8178
atg gtt gag gac tgg ctg cac cga ggg cca agg gaa ttt tgc gtg aag Met Val Glu Asp Trp Leu His Arg Gly Pro Arg Glu Phe Cys Val Lys 2695 2700 2705 2710	8226
gtg ctc tgc ccc tac atg ccg aaa gtc ata gag aag atg gag ctg ctc Val Leu Cys Pro Tyr Met Pro Lys Val Ile Glu Lys Met Glu Leu Leu 2715 2720 2725	8274

caa cgc cgg tat ggg ggg gga ctg gtc aga aac cca ctc tca cgg aat Gln Arg Arg Tyr Gly Gly Gly Leu Val Arg Asn Pro Leu Ser Arg Asn 2730 2735 2740	8322
tcc acg cac gag atg tat tgg gtg agt cga gct tca ggc aat gtg gta Ser Thr His Glu Met Tyr Trp Val Ser Arg Ala Ser Gly Asn Val Val 2745 2750 2755	8370
cat tca gtg aat atg acc agc cag gtg ctc cta gga aga atg gaa aaa His Ser Val Asn Met Thr Ser Gln Val Leu Leu Gly Arg Met Glu Lys 2760 2765 2770	8418
agg acc tgg aag gga ccc caa tac gag gaa gat gta aac ttg gga agc Arg Thr Trp Lys Gly Pro Gln Tyr Glu Asp Val Asn Leu Gly Ser 2775 2780 2785 2790	8466
gga acc agg gcg gtg gga aaa ccc ctg ctc aac tca gac acc agt aaa Gly Thr Arg Ala Val Gly Lys Pro Leu Leu Asn Ser Asp Thr Ser Lys 2795 2800 2805	8514
atc aac aac agg att gaa cga ctc agg cgt gag tac agt tcg acg tgg Ile Asn Asn Arg Ile Glu Arg Leu Arg Arg Glu Tyr Ser Ser Thr Trp 2810 2815 2820	8562
cac cac gat gag aac cac cca tat aga acc tgg aac tat cac ggc agt His His Asp Glu Asn His Pro Tyr Arg Thr Trp Asn Tyr His Gly Ser 2825 2830 2835	8610
tat gat gtg aag ccc aca ggc tcc gcc agt tcg ctg gtc aat gga gtg Tyr Asp Val Lys Pro Thr Gly Ser Ala Ser Ser Leu Val Asn Gly Val 2840 2845 2850	8658
gtc agg ctc ctc tca aaa cca tgg gac acc atc acg aat gtt acc acc Val Arg Leu Leu Ser Lys Pro Trp Asp Thr Ile Thr Asn Val Thr Thr 2855 2860 2865 2870	8706
atg gcc atg act gac act act ccc ttc ggg cag cag cga gtg ttc aaa Met Ala Met Thr Asp Thr Thr Pro Phe Gly Gln Gln Arg Val Phe Lys 2875 2880 2885	8754
gag aag gtg gac acg aaa gct cct gaa ccg cca gaa gga gcg aag tac Glu Lys Val Asp Thr Lys Ala Pro Glu Pro Pro Glu Gly Ala Lys Tyr 2890 2895 2900	8802
gtg ctc aac gag acc acc aac tgg ttg tgg gcg ttt ttg gcc aga gaa Val Leu Asn Glu Thr Thr Asn Trp Leu Trp Ala Phe Leu Ala Arg Glu 2905 2910 2915	8850
aaa cgt ccc aga atg tgc tct cga gag gaa ttc ata aga aag gtc aac Lys Arg Pro Arg Met Cys Ser Arg Glu Glu Phe Ile Arg Lys Val Asn 2920 2925 2930	8898
agc aat gca gct ttg ggt gcc atg ttt gaa gag cag aat caa tgg agg Ser Asn Ala Ala Leu Gly Ala Met Phe Glu Glu Gln Asn Gln Trp Arg 2935 2940 2945 2950	8946
agc gcc aga gaa gca gtt gaa gat cca aaa ttt tgg gag atg gtg gat Ser Ala Arg Glu Ala Val Glu Asp Pro Lys Phe Trp Glu Met Val Asp 2955 2960 2965	8994

gag gag cgc gag gca cat ctg cgg ggg gaa tgt cac act tgc att tac Glu Glu Arg Glu Ala His Leu Arg Gly Glu Cys His Thr Cys Ile Tyr 2970 2975 2980	9042
aac atg atg gga aag aga gag aaa aaa ccc gga gag ttc gga aag gcc Asn Met Met Gly Lys Arg Glu Lys Lys Pro Gly Glu Phe Gly Lys Ala 2985 2990 2995	9090
aag gga agc aga gcc att tgg ttc atg tgg ctc gga gct cgc ttt ctg Lys Gly Ser Arg Ala Ile Trp Phe Met Trp Leu Gly Ala Arg Phe Leu 3000 3005 3010	9138
gag ttc gag gct ctg ggt ttt ctc aat gaa gac cac tgg ctt gga aga Glu Phe Glu Ala Leu Gly Phe Leu Asn Glu Asp His Trp Leu Gly Arg 3015 3020 3025 3030	9186
aag aac tca gga gga ggt gtc gag ggc ttg ggc ctc caa aaa ctg ggt Lys Asn Ser Gly Gly Gly Val Glu Gly Leu Gly Leu Gln Lys Leu Gly 3035 3040 3045	9234
tac atc ctg cgt gaa gtt ggc acc cgg cct ggg ggc aag atc tat gct Tyr Ile Leu Arg Glu Val Gly Thr Arg Pro Gly Gly Lys Ile Tyr Ala 3050 3055 3060	9282
gat gac aca gct ggc tgg gac acc cgc atc acg aga gct gac ttg gaa Asp Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Arg Ala Asp Leu Glu 3065 3070 3075	9330
aat gaa gct aag gtg ctt gag ctg ctt gat ggg gaa cat cgg cgt ctt Asn Glu Ala Lys Val Leu Glu Leu Leu Asp Gly Glu His Arg Arg Leu 3080 3085 3090	9378
gcc agg gcc atc att gag ctc acc tat cgt cac aaa gtt gtg aaa gtg Ala Arg Ala Ile Ile Glu Leu Thr Tyr Arg His Lys Val Val Lys Val 3095 3100 3105 3110	9426
atg cgc ccg gct gct gat gga aga acc gtc atg gat gtt atc tcc aga Met Arg Pro Ala Ala Asp Gly Arg Thr Val Met Asp Val Ile Ser Arg 3115 3120 3125	9474
gaa gat cag agg ggg agt gga caa gtt gtc acc tac gcc cta aac act Glu Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala Leu Asn Thr 3130 3135 3140	9522
ttc acc aac ctg gcc gtc cag ctg gtg agg atg atg gaa ggg gaa gga Phe Thr Asn Leu Ala Val Gln Leu Val Arg Met Met Glu Gly Glu Gly 3145 3150 3155	9570
gtg att ggc cca gat gat gtg gag aaa ctc aca aaa ggg aaa gga ccc Val Ile Gly Pro Asp Asp Val Glu Lys Leu Thr Lys Gly Lys Gly Pro 3160 3165 3170	9618
aaa gtc agg acc tgg ctg ttt gag aat ggg gaa gaa aga ctc agc cgc Lys Val Arg Thr Trp Leu Phe Glu Asn Gly Glu Glu Arg Leu Ser Arg 3175 3180 3185 3190	9666
atg gct gtc agt gga gat gac tgt gtg gta aag ccc ctg gac gat cgc Met Ala Val Ser Gly Asp Asp Cys Val Val Lys Pro Leu Asp Asp Arg 3195 3200 3205	9714

ttt gcc acc tcg ctc cac ttc ctc aat gct atg tca aag gtt cgc aaa 9762
 Phe Ala Thr Ser Leu His Phe Leu Asn Ala Met Ser Lys Val Arg Lys
 3210 3215 3220

gac atc caa gag tgg aaa ccg tca act gga tgg tat gat tgg cag cag 9810
 Asp Ile Gln Glu Trp Lys Pro Ser Thr Gly Trp Tyr Asp Trp Gln Gln
 3225 3230 3235

gtt cca ttt tgc tca aac cat ttc act gaa ttg atc atg aaa gat gga 9858
 Val Pro Phe Cys Ser Asn His Phe Thr Glu Leu Ile Met Lys Asp Gly
 3240 3245 3250

aga aca ctg gtg gtt cca tgc cga gga cag gat gaa ttg gta ggc aga 9906
 Arg Thr Leu Val Val Pro Cys Arg Gly Gln Asp Glu Leu Val Gly Arg
 3255 3260 3265 3270

gct cgc ata tct cca ggg gcc gga tgg aac gtc cgc gac act gct tgt 9954
 Ala Arg Ile Ser Pro Gly Ala Gly Trp Asn Val Arg Asp Thr Ala Cys
 3275 3280 3285

ctg gct aag tct tat gcc cag atg tgg ctg ctt ctg tac ttc cac aga 10002
 Leu Ala Lys Ser Tyr Ala Gln Met Trp Leu Leu Leu Tyr Phe His Arg
 3290 3295 3300

aga gac ctg cgg ctc atg gcc aac gcc att tgc tcc gct gtc cct gtg 10050
 Arg Asp Leu Arg Leu Met Ala Asn Ala Ile Cys Ser Ala Val Pro Val
 3305 3310 3315

aat tgg gtc cct acc gga aga acc acg tgg tcc atc cat gca gga gga 10098
 Asn Trp Val Pro Thr Gly Arg Thr Thr Trp Ser Ile His Ala Gly Gly
 3320 3325 3330

gag tgg atg aca aca gag gac atg ttg gag gtc tgg aac cgt gtt tgg 10146
 Glu Trp Met Thr Thr Glu Asp Met Leu Glu Val Trp Asn Arg Val Trp
 3335 3340 3345 3350

ata gag gag aat gaa tgg atg gaa gac aaa acc cca gtg gag aaa tgg 10194
 Ile Glu Glu Asn Glu Trp Met Glu Asp Lys Thr Pro Val Glu Lys Trp
 3355 3360 3365

agt gac gtc cca tat tca gga aaa cga gag gac atc tgg tgt ggc agc 10242
 Ser Asp Val Pro Tyr Ser Gly Lys Arg Glu Asp Ile Trp Cys Gly Ser
 3370 3375 3380

ctg att ggc aca aga gcc cga gcc acg tgg gca gaa aac atc cag gtg 10290
 Leu Ile Gly Thr Arg Ala Arg Ala Thr Trp Ala Glu Asn Ile Gln Val
 3385 3390 3395

gct atc aac caa gtc aga gca atc atc gga gat gag aag tat gtg gat 10338
 Ala Ile Asn Gln Val Arg Ala Ile Ile Gly Asp Glu Lys Tyr Val Asp
 3400 3405 3410

tac atg agt tca cta aag aga tat gaa gac aca act ttg gtt gag gac 10386
 Tyr Met Ser Ser Leu Lys Arg Tyr Glu Asp Thr Thr Leu Val Glu Asp
 3415 3420 3425 3430

aca gta ctg tagatatatta atcaattgta aatagacaat ataagtatgc 10435
 Thr Val Leu

ataaaagtgt agttttatag tagtatattag tggtgttagt gtaaatagtt aagaaaattt 10495

tgaggagaaa gtcaggccgg gaagttcccg ccaccggaag ttgagtagac ggtgctgcct 10555
 gcgactcaac cccaggagga ctgggtgaac aaagccgcga agtgatccat gtaagccctc 10615
 agaaccgtct cggaaggagg accccacatg ttgtaacttc aaagcccaat gtcagaccac 10675
 gctacggcgt gctactctgc ggagagtgca gtctgcgata gtgcccagg aggactgggt 10735
 taacaaaggc aaaccaacgc cccacgcggc cctagccccg gtaatggtgt taaccagggc 10795
 gaaaggacta gaggttagag gagacccgcg ggtttaaagt gcacggccca gcctgactga 10855
 agctgtaggt caggggaagg actagagggt agtggagacc ccgtgccaca aaacaccaca 10915
 acaaaacagc atattgacac ctgggataga ctaggagatc ttctgctctg cacaaccagc 10975
 cacacggcac agtgcgccga caatggtggc tggtggtgcg agaacacagg atct 11029

<210> 2

<211> 3433

<212> PRT

<213> Flavivirus sp.

<400> 2

Met	Ser	Lys	Lys	Pro	Gly	Gly	Pro	Gly	Lys	Ser	Arg	Ala	Val	Asn	Met
1				5					10					15	
Leu	Lys	Arg	Gly	Met	Pro	Arg	Val	Leu	Ser	Leu	Ile	Gly	Leu	Lys	Arg
			20					25					30		
Ala	Met	Leu	Ser	Leu	Ile	Asp	Gly	Lys	Gly	Pro	Ile	Arg	Phe	Val	Leu
		35					40					45			
Ala	Leu	Leu	Ala	Phe	Phe	Arg	Phe	Thr	Ala	Ile	Ala	Pro	Thr	Arg	Ala
	50					55					60				
Val	Leu	Asp	Arg	Trp	Arg	Gly	Val	Asn	Lys	Gln	Thr	Ala	Met	Lys	His
65					70					75					80
Leu	Leu	Ser	Phe	Lys	Lys	Glu	Leu	Gly	Thr	Leu	Thr	Ser	Ala	Ile	Asn
				85					90					95	
Arg	Arg	Ser	Ser	Lys	Gln	Lys	Lys	Arg	Gly	Gly	Lys	Thr	Gly	Ile	Ala
			100					105					110		
Val	Met	Ile	Gly	Leu	Ile	Ala	Ser	Val	Gly	Ala	Val	Thr	Leu	Ser	Asn
		115					120					125			
Phe	Gln	Gly	Lys	Val	Met	Met	Thr	Val	Asn	Ala	Thr	Asp	Val	Thr	Asp
		130					135					140			
Val	Ile	Thr	Ile	Pro	Thr	Ala	Ala	Gly	Lys	Asn	Leu	Cys	Ile	Val	Arg
145					150					155					160
Ala	Met	Asp	Val	Gly	Tyr	Met	Cys	Asp	Asp	Thr	Ile	Thr	Tyr	Glu	Cys
				165					170					175	
Pro	Val	Leu	Ser	Ala	Gly	Asn	Asp	Pro	Glu	Asp	Ile	Asp	Cys	Trp	Cys
				180					185					190	

17/42

Thr Lys Ser Ala Val Tyr Val Arg Tyr Gly Arg Cys Thr Lys Thr Arg
 195 200 205
 His Ser Arg Arg Ser Arg Arg Ser Leu Thr Val Gln Thr His Gly Glu
 210 215 220
 Ser Thr Leu Ala Asn Lys Lys Gly Ala Trp Met Asp Ser Thr Lys Ala
 225 230 235 240
 Thr Arg Tyr Leu Val Lys Thr Glu Ser Trp Ile Leu Arg Asn Pro Gly
 245 250 255
 Tyr Ala Leu Val Ala Ala Val Ile Gly Trp Met Leu Gly Ser Asn Thr
 260 265 270
 Met Gln Arg Val Val Phe Val Val Leu Leu Leu Leu Val Ala Pro Ala
 275 280 285
 Tyr Ser Phe Asn Cys Leu Gly Met Ser Asn Arg Asp Phe Leu Glu Gly
 290 295 300
 Val Ser Gly Ala Thr Trp Val Asp Leu Val Leu Glu Gly Asp Ser Cys
 305 310 315 320
 Val Thr Ile Met Ser Lys Asp Lys Pro Thr Ile Asp Val Lys Met Met
 325 330 335
 Asn Met Glu Ala Ala Asn Leu Ala Glu Val Arg Ser Tyr Cys Tyr Leu
 340 345 350
 Ala Thr Val Ser Asp Leu Ser Thr Lys Ala Ala Cys Pro Thr Met Gly
 355 360 365
 Glu Ala His Asn Asp Lys Arg Ala Asp Pro Ala Phe Val Cys Arg Gln
 370 375 380
 Gly Val Val Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys
 385 390 395 400
 Gly Ser Ile Asp Thr Cys Ala Lys Phe Ala Cys Ser Thr Lys Ala Ile
 405 410 415
 Gly Arg Thr Ile Leu Lys Glu Asn Ile Lys Tyr Glu Val Ala Ile Phe
 420 425 430
 Val His Gly Pro Thr Thr Val Glu Ser His Gly Asn Tyr Ser Thr Gln
 435 440 445
 Val Gly Ala Thr Gln Ala Gly Arg Phe Ser Ile Thr Pro Ala Ala Pro
 450 455 460
 Ser Tyr Thr Leu Lys Leu Gly Glu Tyr Gly Glu Val Thr Val Asp Cys
 465 470 475 480
 Glu Pro Arg Ser Gly Ile Asp Thr Asn Ala Tyr Tyr Val Met Thr Val
 485 490 495
 Gly Thr Lys Thr Phe Leu Val His Arg Glu Trp Phe Met Asp Leu Asn
 500 505 510

Leu Pro Trp Ser Ser Ala Gly Ser Thr Val Trp Arg Asn Arg Glu Thr
 515 520 525
 Leu Met Glu Phe Glu Glu Pro His Ala Thr Lys Gln Ser Val Ile Ala
 530 535 540
 Leu Gly Ser Gln Glu Gly Ala Leu His Gln Ala Leu Ala Gly Ala Ile
 545 550 555 560
 Pro Val Glu Phe Ser Ser Asn Thr Val Lys Leu Thr Ser Gly His Leu
 565 570 575
 Lys Cys Arg Val Lys Met Glu Lys Leu Gln Leu Lys Gly Thr Thr Tyr
 580 585 590
 Gly Val Cys Ser Lys Ala Phe Lys Phe Leu Gly Thr Pro Ala Asp Thr
 595 600 605
 Gly His Gly Thr Val Val Leu Glu Leu Gln Tyr Thr Gly Thr Asp Gly
 610 615 620
 Pro Cys Lys Val Pro Ile Ser Ser Val Ala Ser Leu Asn Asp Leu Thr
 625 630 635 640
 Pro Val Gly Arg Leu Val Thr Val Asn Pro Phe Val Ser Val Ala Thr
 645 650 655
 Ala Asn Ala Lys Val Leu Ile Glu Leu Glu Pro Pro Phe Gly Asp Ser
 660 665 670
 Tyr Ile Val Val Gly Arg Gly Glu Gln Gln Ile Asn His His Trp His
 675 680 685
 Lys Ser Gly Ser Ser Ile Gly Lys Ala Phe Thr Thr Thr Leu Lys Gly
 690 695 700
 Ala Gln Arg Leu Ala Ala Leu Gly Asp Thr Ala Trp Asp Phe Gly Ser
 705 710 715 720
 Val Gly Gly Val Phe Thr Ser Val Gly Lys Ala Val His Gln Val Phe
 725 730 735
 Gly Gly Ala Phe Arg Ser Leu Phe Gly Gly Met Ser Trp Ile Thr Gln
 740 745 750
 Gly Leu Leu Gly Ala Leu Leu Leu Trp Met Gly Ile Asn Ala Arg Asp
 755 760 765
 Arg Ser Ile Ala Leu Thr Phe Leu Ala Val Gly Gly Val Leu Leu Phe
 770 775 780
 Leu Ser Val Asn Val His Ala Asp Thr Gly Cys Ala Ile Asp Ile Ser
 785 790 795 800
 Arg Gln Glu Leu Arg Cys Gly Asn Gly Val Phe Ile His Asn Asp Val
 805 810 815
 Glu Ala Trp Met Asp Arg Tyr Lys Tyr Tyr Pro Glu Thr Pro Gln Gly
 820 825 830

19/42

Leu Ala Lys Ile Ile Gln Lys Ala His Lys Glu Gly Val Cys Gly Leu
 835 840 845

Arg Ser Val Ser Arg Leu Glu His Gln Met Trp Glu Ala Val Lys Asp
 850 855 860

Glu Leu Asn Thr Leu Leu Lys Glu Asn Gly Val Asp Leu Ser Val Val
 865 870 875 880

Val Glu Lys Gln Glu Gly Met Tyr Lys Ser Ala Pro Lys Arg Leu Thr
 885 890 895

Ala Thr Thr Glu Lys Leu Glu Ile Gly Trp Lys Ala Trp Gly Lys Ser
 900 905 910

Ile Leu Phe Ala Pro Glu Leu Ala Asn Asn Thr Phe Val Val Asp Gly
 915 920 925

Pro Glu Thr Lys Glu Cys Pro Thr Gln Asn Arg Ala Trp Asn Ser Leu
 930 935 940

Glu Val Glu Asp Phe Gly Phe Gly Leu Thr Ser Thr Arg Met Phe Leu
 945 950 955 960

Lys Val Arg Glu Ser Asn Thr Thr Glu Cys Asp Ser Lys Ile Ile Gly
 965 970 975

Thr Ala Val Lys Asn Asn Leu Ala Ile His Ser Asp Leu Ser Tyr Trp
 980 985 990

Ile Glu Ser Arg Leu Asn Asp Thr Trp Lys Leu Glu Arg Ala Val Leu
 995 1000 1005

Gly Glu Val Lys Ser Cys Thr Trp Pro Glu Thr His Thr Leu Trp Gly
 1010 1015 1020

Asp Gly Ile Leu Glu Ser Asp Leu Ile Ile Pro Val Thr Leu Ala Gly
 1025 1030 1035 1040

Pro Arg Ser Asn His Asn Arg Arg Pro Gly Tyr Lys Thr Gln Asn Gln
 1045 1050 1055

Gly Pro Trp Asp Glu Gly Arg Val Glu Ile Asp Phe Asp Tyr Cys Pro
 1060 1065 1070

Gly Thr Thr Val Thr Leu Ser Glu Ser Cys Gly His Arg Gly Pro Ala
 1075 1080 1085

Thr Arg Thr Thr Thr Glu Ser Gly Lys Leu Ile Thr Asp Trp Cys Cys
 1090 1095 1100

Arg Ser Cys Thr Leu Pro Pro Leu Arg Tyr Gln Thr Asp Ser Gly Cys
 1105 1110 1115 1120

Trp Tyr Gly Met Glu Ile Arg Pro Gln Arg His Asp Glu Lys Thr Leu
 1125 1130 1135

Val Gln Ser Gln Val Asn Ala Tyr Asn Ala Asp Met Ile Asp Pro Phe
 1140 1145 1150

Gln Leu Gly Leu Leu Val Val Phe Leu Ala Thr Gln Glu Val Leu Arg
1155 1160 1165

Lys Arg Trp Thr Ala Lys Ile Ser Met Pro Ala Ile Leu Ile Ala Leu
1170 1175 1180

Leu Val Leu Val Phe Gly Gly Ile Thr Tyr Thr Asp Val Leu Arg Tyr
185 1190 1195 1200

Val Ile Leu Val Gly Ala Ala Phe Ala Glu Ser Asn Ser Gly Gly Asp
1205 1210 1215

Val Val His Leu Ala Leu Met Ala Thr Phe Lys Ile Gln Pro Val Phe
1220 1225 1230

Met Val Ala Ser Phe Leu Lys Ala Arg Trp Thr Asn Gln Glu Asn Ile
1235 1240 1245

Leu Leu Met Leu Ala Ala Val Phe Phe Gln Met Ala Tyr His Asp Ala
1250 1255 1260

Arg Gln Ile Leu Leu Trp Glu Ile Pro Asp Val Leu Asn Ser Leu Ala
265 1270 1275 1280

Val Ala Trp Met Ile Leu Arg Ala Ile Thr Phe Thr Thr Thr Ser Asn
1285 1290 1295

Val Val Val Pro Leu Leu Ala Leu Leu Thr Pro Arg Leu Arg Cys Leu
1300 1305 1310

Asn Leu Asp Val Tyr Arg Ile Leu Leu Leu Met Val Gly Ile Gly Ser
1315 1320 1325

Leu Ile Arg Glu Lys Arg Ser Ala Ala Ala Lys Lys Lys Gly Ala Ser
1330 1335 1340

Leu Leu Cys Leu Ala Leu Ala Ser Thr Gly Leu Phe Asn Pro Met Ile
345 1350 1355 1360

Leu Ala Ala Gly Leu Ile Ala Cys Asp Pro Asn Arg Lys Arg Gly Trp
1365 1370 1375

Pro Ala Thr Glu Val Met Thr Ala Val Gly Leu Met Phe Ala Ile Val
1380 1385 1390

Gly Gly Leu Ala Glu Leu Asp Ile Asp Ser Met Ala Ile Pro Met Thr
1395 1400 1405

Ile Ala Gly Leu Met Phe Ala Ala Phe Val Ile Ser Gly Lys Ser Thr
1410 1415 1420

Asp Met Trp Ile Glu Arg Thr Ala Asp Ile Ser Trp Glu Ser Asp Ala
425 1430 1435 1440

Glu Ile Thr Gly Ser Ser Glu Arg Val Asp Val Arg Leu Asp Asp Gly
1445 1450 1455

Glu Asn Phe Gln Leu Met Asn Asp Pro Gly Ala Pro Trp Lys Ile Trp
1460 1465 1470

Met Leu Arg Met Val Cys Leu Ala Ile Ser Ala Tyr Thr Pro Trp Ala
 1475 1480 1485
 Ile Leu Pro Ser Val Val Gly Phe Trp Ile Thr Leu Gln Tyr Thr Lys
 1490 1495 1500
 Arg Gly Gly Val Leu Trp Asp Thr Pro Ser Pro Lys Glu Tyr Lys Lys
 505 1510 1515 1520
 Gly Asp Thr Thr Thr Gly Val Tyr Arg Ile Met Thr Arg Gly Leu Leu
 1525 1530 1535
 Gly Ser Tyr Gln Ala Gly Ala Gly Val Met Val Glu Gly Val Phe His
 1540 1545 1550
 Thr Leu Trp His Thr Thr Lys Gly Ala Ala Leu Met Ser Gly Glu Gly
 1555 1560 1565
 Arg Leu Asp Pro Tyr Trp Gly Ser Val Lys Glu Asp Arg Leu Cys Tyr
 1570 1575 1580
 Gly Gly Pro Trp Lys Leu Gln His Lys Trp Asn Gly Gln Asp Glu Val
 585 1590 1595 1600
 Gln Met Ile Val Val Glu Pro Gly Lys Asn Val Lys Asn Val Gln Thr
 1605 1610 1615
 Lys Pro Gly Val Phe Lys Thr Pro Glu Gly Glu Ile Gly Ala Val Thr
 1620 1625 1630
 Leu Asp Phe Pro Thr Gly Thr Ser Gly Ser Pro Ile Val Asp Lys Asn
 1635 1640 1645
 Gly Asp Val Ile Gly Leu Tyr Gly Asn Gly Val Ile Met Pro Asn Gly
 1650 1655 1660
 Ser Tyr Ile Ser Ala Ile Val Gln Gly Glu Arg Met Asp Glu Pro Ile
 665 1670 1675 1680
 Pro Ala Gly Phe Glu Pro Glu Met Leu Arg Lys Lys Gln Ile Thr Val
 1685 1690 1695
 Leu Asp Leu His Pro Gly Ala Gly Lys Thr Arg Arg Ile Leu Pro Gln
 1700 1705 1710
 Ile Ile Lys Glu Ala Ile Asn Arg Arg Leu Arg Thr Ala Val Leu Ala
 1715 1720 1725
 Pro Thr Arg Val Val Ala Ala Glu Met Ala Glu Ala Leu Arg Gly Leu
 1730 1735 1740
 Pro Ile Arg Tyr Gln Thr Ser Ala Val Pro Arg Glu His Asn Gly Asn
 745 1750 1755 1760
 Glu Ile Val Asp Val Met Cys His Ala Thr Leu Thr His Arg Leu Met
 1765 1770 1775
 Ser Pro His Arg Val Pro Asn Tyr Asn Leu Phe Val Met Asp Glu Ala
 1780 1785 1790

22/42

His Phe Thr Asp Pro Ala Ser Ile Ala Ala Arg Gly Tyr Ile Ser Thr
1795 1800 1805

Lys Val Glu Leu Gly Glu Ala Ala Ala Ile Phe Met Thr Ala Thr Pro
1810 1815 1820

Pro Gly Thr Ser Asp Pro Phe Pro Glu Ser Asn Ser Pro Ile Ser Asp
825 1830 1835 1840

Leu Gln Thr Glu Ile Pro Asp Arg Ala Trp Asn Ser Gly Tyr Glu Trp
1845 1850 1855

Ile Thr Glu Tyr Thr Gly Lys Thr Val Trp Phe Val Pro Ser Val Lys
1860 1865 1870

Met Gly Asn Glu Ile Ala Leu Cys Leu Gln Arg Ala Gly Lys Lys Val
1875 1880 1885

Val Gln Leu Asn Arg Lys Ser Tyr Glu Thr Glu Tyr Pro Lys Cys Lys
1890 1895 1900

Asn Asp Asp Trp Asp Phe Val Ile Thr Thr Asp Ile Ser Glu Met Gly
905 1910 1915 1920

Ala Asn Phe Lys Ala Ser Arg Val Ile Asp Ser Arg Lys Ser Val Lys
1925 1930 1935

Pro Thr Ile Ile Thr Glu Gly Glu Gly Arg Val Ile Leu Gly Glu Pro
1940 1945 1950

Ser Ala Val Thr Ala Ala Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly
1955 1960 1965

Arg Asn Pro Ser Gln Val Gly Asp Glu Tyr Cys Tyr Gly Gly His Thr
1970 1975 1980

Asn Glu Asp Asp Ser Asn Phe Ala His Trp Thr Glu Ala Arg Ile Met
985 1990 1995 2000

Pro Asp Asn Ile Asn Met Pro Asn Gly Leu Ile Ala Gln Phe Tyr Gln
2005 2010 2015

Pro Glu Arg Glu Lys Val Tyr Thr Met Glu Gly Glu Tyr Arg Leu Arg
2020 2025 2030

Gly Glu Glu Arg Lys Asn Phe Leu Glu Leu Leu Arg Thr Ala Asp Leu
2035 2040 2045

Pro Val Trp Leu Ala Tyr Lys Val Ala Ala Ala Gly Val Ser Tyr His
2050 2055 2060

Asp Arg Arg Trp Cys Phe Asp Gly Pro Arg Thr Asn Thr Ile Leu Glu
065 2070 2075 2080

Asp Asn Asn Glu Val Glu Val Ile Thr Lys Leu Gly Glu Arg Lys Ile
2085 2090 2095

Leu Arg Pro Arg Trp Ile Asp Ala Arg Val Tyr Ser Asp His Gln Ala
2100 2105 2110

23/42

Leu Lys Ala Phe Lys Asp Phe Ala Ser Gly Lys Arg Ser Gln Ile Gly
2115 2120 2125

Leu Ile Glu Val Leu Gly Lys Met Pro Glu His Phe Met Gly Lys Thr
2130 2135 2140

Trp Glu Ala Leu Asp Thr Met Tyr Val Val Ala Thr Ala Glu Lys Gly
145 2150 2155 2160

Gly Arg Ala His Arg Met Ala Leu Glu Glu Leu Pro Asp Ala Leu Gln
2165 2170 2175

Thr Ile Ala Leu Ile Ala Leu Leu Ser Val Met Thr Met Gly Val Phe
2180 2185 2190

Phe Leu Leu Met Gln Arg Lys Gly Ile Gly Lys Ile Gly Leu Gly Gly
2195 2200 2205

Ala Val Leu Gly Val Ala Thr Phe Phe Cys Trp Met Ala Glu Val Pro
2210 2215 2220

Gly Thr Lys Ile Ala Gly Met Leu Leu Leu Ser Leu Leu Leu Met Ile
225 2230 2235 2240

Val Leu Ile Pro Glu Pro Glu Lys Gln Arg Ser Gln Thr Asp Asn Gln
2245 2250 2255

Leu Ala Val Phe Leu Ile Cys Val Met Thr Leu Val Ser Ala Val Ala
2260 2265 2270

Ala Asn Glu Met Gly Trp Leu Asp Lys Thr Lys Ser Asp Ile Ser Ser
2275 2280 2285

Leu Phe Gly Gln Arg Ile Glu Val Lys Glu Asn Phe Ser Met Gly Glu
2290 2295 2300

Phe Leu Leu Asp Leu Arg Pro Ala Thr Ala Trp Ser Leu Tyr Ala Val
305 2310 2315 2320

Thr Thr Ala Val Leu Thr Pro Leu Leu Lys His Leu Ile Thr Ser Asp
2325 2330 2335

Tyr Ile Asn Thr Ser Leu Thr Ser Ile Asn Val Gln Ala Ser Ala Leu
2340 2345 2350

Phe Thr Leu Ala Arg Gly Phe Pro Phe Val Asp Val Gly Val Ser Ala
2355 2360 2365

Leu Leu Leu Ala Ala Gly Cys Trp Gly Gln Val Thr Leu Thr Val Thr
2370 2375 2380

Val Thr Ala Ala Thr Leu Leu Phe Cys His Tyr Ala Tyr Met Val Pro
385 2390 2395 2400

Gly Trp Gln Ala Glu Ala Met Arg Ser Ala Gln Arg Arg Thr Ala Ala
2405 2410 2415

Gly Ile Met Lys Asn Ala Val Val Asp Gly Ile Val Ala Thr Asp Val
2420 2425 2430

24/42

Pro Glu Leu Glu Arg Thr Thr Pro Ile Met Gln Lys Lys Val Gly Gln
2435 2440 2445

Ile Met Leu Ile Leu Val Ser Leu Ala Ala Val Val Val Asn Pro Ser
2450 2455 2460

Val Lys Thr Val Arg Glu Ala Gly Ile Leu Ile Thr Ala Ala Ala Val
465 2470 2475 2480

Thr Leu Trp Glu Asn Gly Ala Ser Ser Val Trp Asn Ala Thr Thr Ala
2485 2490 2495

Ile Gly Leu Cys His Ile Met Arg Gly Gly Trp Leu Ser Cys Leu Ser
2500 2505 2510

Ile Thr Trp Thr Leu Ile Lys Asn Met Glu Lys Pro Gly Leu Lys Arg
2515 2520 2525

Gly Gly Ala Lys Gly Arg Thr Leu Gly Glu Val Trp Lys Glu Arg Leu
2530 2535 2540

Asn Gln Met Thr Lys Glu Glu Phe Thr Arg Tyr Arg Lys Glu Ala Ile
545 2550 2555 2560

Ile Glu Val Asp Arg Ser Ala Ala Lys His Ala Arg Lys Glu Gly Asn
2565 2570 2575

Val Thr Gly Gly His Ser Val Ser Arg Gly Thr Ala Lys Leu Arg Trp
2580 2585 2590

Leu Val Glu Arg Arg Phe Leu Glu Pro Val Gly Lys Val Ile Asp Leu
2595 2600 2605

Gly Cys Gly Arg Gly Gly Trp Cys Tyr Tyr Met Ala Thr Gln Lys Arg
2610 2615 2620

Val Gln Glu Val Arg Gly Tyr Thr Lys Gly Gly Pro Gly His Glu Glu
625 2630 2635 2640

Pro Gln Leu Val Gln Ser Tyr Gly Trp Asn Ile Val Thr Met Lys Ser
2645 2650 2655

Gly Val Asp Val Phe Tyr Arg Pro Ser Glu Cys Cys Asp Thr Leu Leu
2660 2665 2670

Cys Asp Ile Gly Glu Ser Ser Ser Ser Ala Glu Val Glu Glu His Arg
2675 2680 2685

Thr Ile Arg Val Leu Glu Met Val Glu Asp Trp Leu His Arg Gly Pro
2690 2695 2700

Arg Glu Phe Cys Val Lys Val Leu Cys Pro Tyr Met Pro Lys Val Ile
705 2710 2715 2720

Glu Lys Met Glu Leu Leu Gln Arg Arg Tyr Gly Gly Gly Leu Val Arg
2725 2730 2735

Asn Pro Leu Ser Arg Asn Ser Thr His Glu Met Tyr Trp Val Ser Arg
2740 2745 2750

Ala Ser Gly Asn Val Val His Ser Val Asn Met Thr Ser Gln Val Leu
2755 2760 2765

Leu Gly Arg Met Glu Lys Arg Thr Trp Lys Gly Pro Gln Tyr Glu Glu
2770 2775 2780

Asp Val Asn Leu Gly Ser Gly Thr Arg Ala Val Gly Lys Pro Leu Leu
785 2790 2795 2800

Asn Ser Asp Thr Ser Lys Ile Asn Asn Arg Ile Glu Arg Leu Arg Arg
2805 2810 2815

Glu Tyr Ser Ser Thr Trp His His Asp Glu Asn His Pro Tyr Arg Thr
2820 2825 2830

Trp Asn Tyr His Gly Ser Tyr Asp Val Lys Pro Thr Gly Ser Ala Ser
2835 2840 2845

Ser Leu Val Asn Gly Val Val Arg Leu Leu Ser Lys Pro Trp Asp Thr
2850 2855 2860

Ile Thr Asn Val Thr Thr Met Ala Met Thr Asp Thr Thr Pro Phe Gly
865 2870 2875 2880

Gln Gln Arg Val Phe Lys Glu Lys Val Asp Thr Lys Ala Pro Glu Pro
2885 2890 2895

Pro Glu Gly Ala Lys Tyr Val Leu Asn Glu Thr Thr Asn Trp Leu Trp
2900 2905 2910

Ala Phe Leu Ala Arg Glu Lys Arg Pro Arg Met Cys Ser Arg Glu Glu
2915 2920 2925

Phe Ile Arg Lys Val Asn Ser Asn Ala Ala Leu Gly Ala Met Phe Glu
2930 2935 2940

Glu Gln Asn Gln Trp Arg Ser Ala Arg Glu Ala Val Glu Asp Pro Lys
945 2950 2955 2960

Phe Trp Glu Met Val Asp Glu Glu Arg Glu Ala His Leu Arg Gly Glu
2965 2970 2975

Cys His Thr Cys Ile Tyr Asn Met Met Gly Lys Arg Glu Lys Lys Pro
2980 2985 2990

Gly Glu Phe Gly Lys Ala Lys Gly Ser Arg Ala Ile Trp Phe Met Trp
2995 3000 3005

Leu Gly Ala Arg Phe Leu Glu Phe Glu Ala Leu Gly Phe Leu Asn Glu
3010 3015 3020

Asp His Trp Leu Gly Arg Lys Asn Ser Gly Gly Gly Val Glu Gly Leu
025 3030 3035 3040

Gly Leu Gln Lys Leu Gly Tyr Ile Leu Arg Glu Val Gly Thr Arg Pro
3045 3050 3055

Gly Gly Lys Ile Tyr Ala Asp Asp Thr Ala Gly Trp Asp Thr Arg Ile
3060 3065 3070

Thr Arg Ala Asp Leu Glu Asn Glu Ala Lys Val Leu Glu Leu Leu Asp
3075 3080 3085

Gly Glu His Arg Arg Leu Ala Arg Ala Ile Ile Glu Leu Thr Tyr Arg
3090 3095 3100

His Lys Val Val Lys Val Met Arg Pro Ala Ala Asp Gly Arg Thr Val
105 3110 3115 3120

Met Asp Val Ile Ser Arg Glu Asp Gln Arg Gly Ser Gly Gln Val Val
3125 3130 3135

Thr Tyr Ala Leu Asn Thr Phe Thr Asn Leu Ala Val Gln Leu Val Arg
3140 3145 3150

Met Met Glu Gly Glu Gly Val Ile Gly Pro Asp Asp Val Glu Lys Leu
3155 3160 3165

Thr Lys Gly Lys Gly Pro Lys Val Arg Thr Trp Leu Phe Glu Asn Gly
3170 3175 3180

Glu Glu Arg Leu Ser Arg Met Ala Val Ser Gly Asp Asp Cys Val Val
185 3190 3195 3200

Lys Pro Leu Asp Asp Arg Phe Ala Thr Ser Leu His Phe Leu Asn Ala
3205 3210 3215

Met Ser Lys Val Arg Lys Asp Ile Gln Glu Trp Lys Pro Ser Thr Gly
3220 3225 3230

Trp Tyr Asp Trp Gln Gln Val Pro Phe Cys Ser Asn His Phe Thr Glu
3235 3240 3245

Leu Ile Met Lys Asp Gly Arg Thr Leu Val Val Pro Cys Arg Gly Gln
3250 3255 3260

Asp Glu Leu Val Gly Arg Ala Arg Ile Ser Pro Gly Ala Gly Trp Asn
265 3270 3275 3280

Val Arg Asp Thr Ala Cys Leu Ala Lys Ser Tyr Ala Gln Met Trp Leu
3285 3290 3295

Leu Leu Tyr Phe His Arg Arg Asp Leu Arg Leu Met Ala Asn Ala Ile
3300 3305 3310

Cys Ser Ala Val Pro Val Asn Trp Val Pro Thr Gly Arg Thr Thr Trp
3315 3320 3325

Ser Ile His Ala Gly Gly Glu Trp Met Thr Thr Glu Asp Met Leu Glu
3330 3335 3340

Val Trp Asn Arg Val Trp Ile Glu Glu Asn Glu Trp Met Glu Asp Lys
345 3350 3355 3360

Thr Pro Val Glu Lys Trp Ser Asp Val Pro Tyr Ser Gly Lys Arg Glu
3365 3370 3375

Asp Ile Trp Cys Gly Ser Leu Ile Gly Thr Arg Ala Arg Ala Thr Trp
3380 3385 3390

27/42

Ala Glu Asn Ile Gln Val Ala Ile Asn Gln Val Arg Ala Ile Ile Gly
3395 3400 3405

Asp Glu Lys Tyr Val Asp Tyr Met Ser Ser Leu Lys Arg Tyr Glu Asp
3410 3415 3420

Thr Thr Leu Val Glu Asp Thr Val Leu
425 3430

<210> 3
<400> 000

<210> 4
<400> 000

<210> 5
<211> 19
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:amorce

<400> 5
gtcagacgct gacctggtg 19

<210> 6
<211> 21
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:amorce

<400> 6
agcttctctt tacacagttg g 21

<210> 7
<211> 19
<212> ADN
<213> Séquence artificielle

<223> Description de la séquence artificielle:amorce

<400> 7
acagtgcagg tgtgtgagc 19

<210> 8
<211> 20
<212> ADN
<213> Séquence artificielle

<

<223> Description de la séquence artificielle:amorce

<400> 8

tcatgtctca gaaaggaaac

20

<210> 9

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: amorce

<400> 9

ctggctgcag aggtaaaagc t

21

<210> 10

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 10

tcaggacagg gtggagtaga gc

22

<210> 11

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 11

gtaccggacc ctcacccctt gt

22

<210> 12

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 12

gctccatcca tccaaccatc ca

22

<210> 13

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 13

gagtttgaag tgaggtgttg c

21

<210> 14

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 14

ttctcttgtg gatgcgttgt g

21

<210> 15

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: amorce

<400> 15

tgggtatatg cggagcgatg c

21

<210> 16

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 16

agtcgaacca acccgctgtc a

21

<210> 17

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 17

actacctcac acggaatcta c

21

<210> 18

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 18

ttctacgcc a atctcatcag t

21

<210> 19

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 19

acacagtgtc catctcaacc a

21

<210> 20

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 20

aaccactggt caaggttctg c

21

<210> 21

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 21

acaaccacgt ccataagtct ctg

23

<210> 22

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 22

gcctttctct tctcagtgt a

21

<210> 23

<211> 1412

<212> ADN

<213> Mus musculus

<220>

<221> CDS

<222> (36)..(1139)

<400> 23

```

ccaggctggg agacccagga agctccagac ttagc atg gag cac gga ctc agg      53
                               Met Glu His Gly Leu Arg
                               1           5

agc atc cca gcc tgg acg ctg gac aag ttc ata gag gat tac ctc ctt      101
Ser Ile Pro Ala Trp Thr Leu Asp Lys Phe Ile Glu Asp Tyr Leu Leu
          10           15           20

ccc gac acc acc ttt ggt gct gat gtc aaa tca gcc gtc aat gtc gtg      149
Pro Asp Thr Thr Phe Gly Ala Asp Val Lys Ser Ala Val Asn Val Val
          25           30           35

tgt gat ttc ctg aag gag aga tgc ttc caa ggt gct gcc cac cca gtg      197
Cys Asp Phe Leu Lys Glu Arg Cys Phe Gln Gly Ala Ala His Pro Val
          40           45           50

agg gtc tcc aag gtg gtg aag ggt ggc tcc tca ggc aaa ggc acc aca      245
Arg Val Ser Lys Val Val Lys Gly Gly Ser Ser Gly Lys Gly Thr Thr
          55           60           65           70

ctc aag ggc agg tca gac gct gac ctg gtg gtg ttc ctt aac aat ctc      293
Leu Lys Gly Arg Ser Asp Ala Asp Leu Val Val Phe Leu Asn Asn Leu
          75           80           85

acc agc ttt gag gat cag tta aac cga cgg gga gag ttc atc aag gaa      341
Thr Ser Phe Glu Asp Gln Leu Asn Arg Arg Gly Glu Phe Ile Lys Glu
          90           95           100

att aag aaa cag ctg tac gag gtt cag cat gag aga cgt ttt aga gtc      389
Ile Lys Lys Gln Leu Tyr Glu Val Gln His Glu Arg Arg Phe Arg Val
          105           110           115

aag ttt gag gtc cag agt tca tgg tgg ccc aac gcc cgg tct ctg agc      437
Lys Phe Glu Val Gln Ser Ser Trp Trp Pro Asn Ala Arg Ser Leu Ser
          120           125           130

ttc aag ctg agc gcc ccc cat ctg cat cag gag gtg gag ttt gat gtg      485
Phe Lys Leu Ser Ala Pro His Leu His Gln Glu Val Glu Phe Asp Val
          135           140           145           150

ctg cca gcc ttt gat gtc ctg ggt cat gtt aat act tcc agc aag cct      533
Leu Pro Ala Phe Asp Val Leu Gly His Val Asn Thr Ser Ser Lys Pro
          155           160           165

gat ccc aga atc tat gcc atc ctc atc gag gaa tgt acc tcc ctg ggg      581
Asp Pro Arg Ile Tyr Ala Ile Leu Ile Glu Glu Cys Thr Ser Leu Gly
          170           175           180

aag gat ggc gag ttc tct acc tgc ttc acg gag ctc cag cgg aac ttc      629
Lys Asp Gly Glu Phe Ser Thr Cys Phe Thr Glu Leu Gln Arg Asn Phe
          185           190           195

ctg aag cag cgc cca acc aag ctg aag agt ctc atc cgc ctg gtc aag      677
Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser Leu Ile Arg Leu Val Lys
          200           205           210

```

cac tgg tac caa ctg tgt aag gag aag ctg ggg aag cca ttg cct cca 725
 His Trp Tyr Gln Leu Cys Lys Glu Lys Leu Gly Lys Pro Leu Pro Pro
 215 220 225 230
 cag tac gcc cta gag ttg ctc act gtc ttt gcc tgg gaa caa ggg aat 773
 Gln Tyr Ala Leu Glu Leu Thr Val Phe Ala Trp Glu Gln Gly Asn
 235 240 245
 gga tgt tat gag ttc aac aca gcc cag ggc ttc cgg acc gtc ttg gaa 821
 Gly Cys Tyr Glu Phe Asn Thr Ala Gln Gly Phe Arg Thr Val Leu Glu
 250 255 260
 ctg gtc atc aat tat cag cat ctt cga atc tac tgg aca aag tat tat 869
 Leu Val Ile Asn Tyr Gln His Leu Arg Ile Tyr Trp Thr Lys Tyr Tyr
 265 270 275
 gac ttt caa cac cag gag gtc tcc aaa tac ctg cac aga cag ctc aga 917
 Asp Phe Gln His Gln Glu Val Ser Lys Tyr Leu His Arg Gln Leu Arg
 280 285 290
 aaa gcc agg cct gtg atc ctg gac cca gct gac cca aca ggg aat gtg 965
 Lys Ala Arg Pro Val Ile Leu Asp Pro Ala Asp Pro Thr Gly Asn Val
 295 300 305 310
 gcc ggt ggg aac cca gag ggc tgg agg cgg ttg gct gaa gag gct gat 1013
 Ala Gly Gly Asn Pro Glu Gly Trp Arg Arg Leu Ala Glu Glu Ala Asp
 315 320 325
 gtg tgg cta tgg tac cca tgt ttt att aaa aag gat ggt tcc cga gtg 1061
 Val Trp Leu Trp Tyr Pro Cys Phe Ile Lys Lys Asp Gly Ser Arg Val
 330 335 340
 agc tcc tgg gat gtg ccg acg gtg gtt cct gta cct ttt gag cag gta 1109
 Ser Ser Trp Asp Val Pro Thr Val Val Pro Val Pro Phe Glu Gln Val
 345 350 355
 gaa gag aac tgg aca tgt atc ctg ctg tga gcacagcagc acctgcccag 1159
 Glu Glu Asn Trp Thr Cys Ile Leu Leu
 360 365
 gagactgctg gtcaggggca tttgctgctc tgctgcaggc ccatgaccca gtgagggagg 1219
 gccccacctg gcatcagact ccgtgcttct gatgcctgcc agccatgttt gactcctgtc 1279
 caatcacagc cagccttcct caacagattc agaaggagag gaaagaacac acgcttggtg 1339
 tccatctgtc cacctgttgg aaggttctgt ctgacaaagt ctgatcaaca ataaaccaca 1399
 gcaggtgccg tca 1412

<210> 24
 <211> 367
 <212> PRT
 <213> Mus musculus

<400> 24
 Met Glu His Gly Leu Arg Ser Ile Pro Ala Trp Thr Leu Asp Lys Phe
 1 5 10 15

Ile Glu Asp Tyr Leu Leu Pro Asp Thr Thr Phe Gly Ala Asp Val Lys
 20 25 30
 Ser Ala Val Asn Val Val Cys Asp Phe Leu Lys Glu Arg Cys Phe Gln
 35 40 45
 Gly Ala Ala His Pro Val Arg Val Ser Lys Val Val Lys Gly Gly Ser
 50 55 60
 Ser Gly Lys Gly Thr Thr Leu Lys Gly Arg Ser Asp Ala Asp Leu Val
 65 70 75 80
 Val Phe Leu Asn Asn Leu Thr Ser Phe Glu Asp Gln Leu Asn Arg Arg
 85 90 95
 Gly Glu Phe Ile Lys Glu Ile Lys Lys Gln Leu Tyr Glu Val Gln His
 100 105 110
 Glu Arg Arg Phe Arg Val Lys Phe Glu Val Gln Ser Ser Trp Trp Pro
 115 120 125
 Asn Ala Arg Ser Leu Ser Phe Lys Leu Ser Ala Pro His Leu His Gln
 130 135 140
 Glu Val Glu Phe Asp Val Leu Pro Ala Phe Asp Val Leu Gly His Val
 145 150 155 160
 Asn Thr Ser Ser Lys Pro Asp Pro Arg Ile Tyr Ala Ile Leu Ile Glu
 165 170 175
 Glu Cys Thr Ser Leu Gly Lys Asp Gly Glu Phe Ser Thr Cys Phe Thr
 180 185 190
 Glu Leu Gln Arg Asn Phe Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser
 195 200 205
 Leu Ile Arg Leu Val Lys His Trp Tyr Gln Leu Cys Lys Glu Lys Leu
 210 215 220
 Gly Lys Pro Leu Pro Pro Gln Tyr Ala Leu Glu Leu Leu Thr Val Phe
 225 230 235 240
 Ala Trp Glu Gln Gly Asn Gly Cys Tyr Glu Phe Asn Thr Ala Gln Gly
 245 250 255
 Phe Arg Thr Val Leu Glu Leu Val Ile Asn Tyr Gln His Leu Arg Ile
 260 265 270
 Tyr Trp Thr Lys Tyr Tyr Asp Phe Gln His Gln Glu Val Ser Lys Tyr
 275 280 285
 Leu His Arg Gln Leu Arg Lys Ala Arg Pro Val Ile Leu Asp Pro Ala
 290 295 300
 Asp Pro Thr Gly Asn Val Ala Gly Gly Asn Pro Glu Gly Trp Arg Arg
 305 310 315 320
 Leu Ala Glu Glu Ala Asp Val Trp Leu Trp Tyr Pro Cys Phe Ile Lys
 325 330 335

Lys Asp Gly Ser Arg Val Ser Ser Trp Asp Val Pro Thr Val Val Pro
 340 345 350

Val Pro Phe Glu Gln Val Glu Glu Asn Trp Thr Cys Ile Leu Leu
 355 360 365

<210> 25
 <211> 1413
 <212> ADN
 <213> Mus musculus

<220>
 <221> CDS
 <222> (36)..(1139)

<400> 25
 ccaggctggg agaccagga agctccagac ttâgc atg gag cac gga ctc agg 53
 Met Glu His Gly Leu Arg
 1 5

agc atc cca gcc tgg acg ctg gac aag ttc ata gag gat tac ctc ctt 101
 Ser Ile Pro Ala Trp Thr Leu Asp Lys Phe Ile Glu Asp Tyr Leu Leu
 10 15 20

ccc gac acc acc ttt ggt gct gat gtc aaa tca gcc gtc aat gtc gtg 149
 Pro Asp Thr Thr Phe Gly Ala Asp Val Lys Ser Ala Val Asn Val Val
 25 30 35

tgt gat ttc ctg aag gag aga tgc ttc caa ggt gct gcc cac cca gtg 197
 Cys Asp Phe Leu Lys Glu Arg Cys Phe Gln Gly Ala Ala His Pro Val
 40 45 50

agg gtc tcc aag gtg gtg aag ggt ggc tcc tca ggc aaa ggc acc aca 245
 Arg Val Ser Lys Val Val Lys Gly Gly Ser Ser Gly Lys Gly Thr Thr
 55 60 65 70

ctc aag ggc agg tca gac gct gac ctg gtg gtg ttc ctt aac aat ctc 293
 Leu Lys Gly Arg Ser Asp Ala Asp Leu Val Val Phe Leu Asn Asn Leu
 75 80 85

acc agc ttt gag gat cag tta aac cga cgg gga gag ttc atc aag gaa 341
 Thr Ser Phe Glu Asp Gln Leu Asn Arg Arg Gly Glu Phe Ile Lys Glu
 90 95 100

att aag aaa cag ctg tac gag gtt cag cat gag aga cgt ttt aga gtc 389
 Ile Lys Lys Gln Leu Tyr Glu Val Gln His Glu Arg Arg Phe Arg Val
 105 110 115

aag ttt gag gtc cag agt tca tgg tgg ccc aac gcc cgg tct ctg agc 437
 Lys Phe Glu Val Gln Ser Ser Trp Trp Pro Asn Ala Arg Ser Leu Ser
 120 125 130

ttc aag ctg agc gcc ccc cat ctg cat cag gag gtg gag ttt gat gtg 485
 Phe Lys Leu Ser Ala Pro His Leu His Gln Glu Val Glu Phe Asp Val
 135 140 145 150

ctg cca gcc ttt gat gtc ctg ggt cat gtt aat act tcc agc aag cct 533
 Leu Pro Ala Phe Asp Val Leu Gly His Val Asn Thr Ser Ser Lys Pro
 155 160 165

gat ccc aga atc tat gcc atc ctc atc gag gaa tgt acc tcc ctg ggg 581
 Asp Pro Arg Ile Tyr Ala Ile Leu Ile Glu Glu Cys Thr Ser Leu Gly
 170 175 180

aag gat ggc gag ttc tct acc tgc ttc acg gag ctc cag cgg aac ttc 629
 Lys Asp Gly Glu Phe Ser Thr Cys Phe Thr Glu Leu Gln Arg Asn Phe
 185 190 195

ctg aag cag cgc cca acc aag ctg aag agt ctc atc cgc ctg gtc aag 677
 Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser Leu Ile Arg Leu Val Lys
 200 205 210

cac tgg tac caa ctg tgt aag gag aag ctg ggg aag cca ttg cct cca 725
 His Trp Tyr Gln Leu Cys Lys Glu Lys Leu Gly Lys Pro Leu Pro Pro
 215 220 225 230

cag tac gcc cta gag ttg ctc act gtc ttt gcc tgg gaa caa ggg aat 773
 Gln Tyr Ala Leu Glu Leu Leu Thr Val Phe Ala Trp Glu Gln Gly Asn
 235 240 245

gga tgt tat gag ttc aac aca gcc cag ggc ttc cgg acc gtc ttg gaa 821
 Gly Cys Tyr Glu Phe Asn Thr Ala Gln Gly Phe Arg Thr Val Leu Glu
 250 255 260

ctg gtc atc aat tat cag cat ctt cga atc tac tgg aca aag tat tat 869
 Leu Val Ile Asn Tyr Gln His Leu Arg Ile Tyr Trp Thr Lys Tyr Tyr
 265 270 275

gac ttt caa cac cag gag gtc tcc aaa tac ctg cac aga cag ctc aga 917
 Asp Phe Gln His Gln Glu Val Ser Lys Tyr Leu His Arg Gln Leu Arg
 280 285 290

aaa gcc agg cct gtg atc ctg gac cca gct gac cca aca ggg aat gtg 965
 Lys Ala Arg Pro Val Ile Leu Asp Pro Ala Asp Pro Thr Gly Asn Val
 295 300 305 310

gcc ggt ggg aac cca gag ggc tgg agg cgg ttg gct gaa gag gct gat 1013
 Ala Gly Gly Asn Pro Glu Gly Trp Arg Arg Leu Ala Glu Glu Ala Asp
 315 320 325

gtg tgg cta tgg tac cca tgt ttt att aaa aag gat ggt tcc cga gtg 1061
 Val Trp Leu Trp Tyr Pro Cys Phe Ile Lys Lys Asp Gly Ser Arg Val
 330 335 340

agc tcc tgg gat gtg ccg acg gtg gtt cct gta cct ttt gag cag gta 1109
 Ser Ser Trp Asp Val Pro Thr Val Val Pro Val Pro Phe Glu Gln Val
 345 350 355

gaa gag aac tgg aca tgt atc ctg ctg tga gcacagcagc acctgccag 1159
 Glu Glu Asn Trp Thr Cys Ile Leu Leu
 360 365

gagactgctg gtcaggggca ttgctgctc tgctgcaggc ccatgacca gtgagggagg 1219

gccccacctg gcatcagact ccgtgcttct gatgcctgcc agccatgttt gactcctgtc 1279

caatcacagc cagccttcct caacagattc agaaggagag gaaagaacac acgcttggtg 1339
 tccatctgtc cacctgttgg aaggttctgt ctgacaaagt ctgatcaaca ataaaccaca 1399
 gcaggtgccc gtca 1413

<210> 26
 <211> 367
 <212> PRT
 <213> Mus musculus

<400> 26
 Met Glu His Gly Leu Arg Ser Ile Pro Ala Trp Thr Leu Asp Lys Phe
 1 5 10 15
 Ile Glu Asp Tyr Leu Leu Pro Asp Thr Thr Phe Gly Ala Asp Val Lys
 20 25 30
 Ser Ala Val Asn Val Val Cys Asp Phe Leu Lys Glu Arg Cys Phe Gln
 35 40 45
 Gly Ala Ala His Pro Val Arg Val Ser Lys Val Val Lys Gly Gly Ser
 50 55 60
 Ser Gly Lys Gly Thr Thr Leu Lys Gly Arg Ser Asp Ala Asp Leu Val
 65 70 75 80
 Val Phe Leu Asn Asn Leu Thr Ser Phe Glu Asp Gln Leu Asn Arg Arg
 85 90 95
 Gly Glu Phe Ile Lys Glu Ile Lys Lys Gln Leu Tyr Glu Val Gln His
 100 105 110
 Glu Arg Arg Phe Arg Val Lys Phe Glu Val Gln Ser Ser Trp Trp Pro
 115 120 125
 Asn Ala Arg Ser Leu Ser Phe Lys Leu Ser Ala Pro His Leu His Gln
 130 135 140
 Glu Val Glu Phe Asp Val Leu Pro Ala Phe Asp Val Leu Gly His Val
 145 150 155 160
 Asn Thr Ser Ser Lys Pro Asp Pro Arg Ile Tyr Ala Ile Leu Ile Glu
 165 170 175
 Glu Cys Thr Ser Leu Gly Lys Asp Gly Glu Phe Ser Thr Cys Phe Thr
 180 185 190
 Glu Leu Gln Arg Asn Phe Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser
 195 200 205
 Leu Ile Arg Leu Val Lys His Trp Tyr Gln Leu Cys Lys Glu Lys Leu
 210 215 220
 Gly Lys Pro Leu Pro Pro Gln Tyr Ala Leu Glu Leu Leu Thr Val Phe
 225 230 235 240
 Ala Trp Glu Gln Gly Asn Gly Cys Tyr Glu Phe Asn Thr Ala Gln Gly
 245 250 255

Phe Arg Thr Val Leu Glu Leu Val Ile Asn Tyr Gln His Leu Arg Ile
 260 265 270
 Tyr Trp Thr Lys Tyr Tyr Asp Phe Gln His Gln Glu Val Ser Lys Tyr
 275 280 285
 Leu His Arg Gln Leu Arg Lys Ala Arg Pro Val Ile Leu Asp Pro Ala
 290 295 300
 Asp Pro Thr Gly Asn Val Ala Gly Gly Asn Pro Glu Gly Trp Arg Arg
 305 310 315 320
 Leu Ala Glu Glu Ala Asp Val Trp Leu Trp Tyr Pro Cys Phe Ile Lys
 325 330 335
 Lys Asp Gly Ser Arg Val Ser Ser Trp Asp Val Pro Thr Val Val Pro
 340 345 350
 Val Pro Phe Glu Gln Val Glu Glu Asn Trp Thr Cys Ile Leu Leu
 355 360 365

<210> 27
 <211> 902
 <212> ADN
 <213> Mus musculus

<220>
 <221> CDS
 <222> (36)..(791)

<400> 27
 acctgctggc tgcagaggta aaagctggac ctagg atg gag cag gat ctg agg 53
 Met Glu Gln Asp Leu Arg
 1 5
 agc atc ccg gcc tcg aag ctt gat aag ttc ata gag aac cat ctc ccg 101
 Ser Ile Pro Ala Ser Lys Leu Asp Lys Phe Ile Glu Asn His Leu Pro
 10 15 20
 gac acc agc ttc tgt gct gac ctc aga gaa gtc ata gat gcc ctg tgt 149
 Asp Thr Ser Phe Cys Ala Asp Leu Arg Glu Val Ile Asp Ala Leu Cys
 25 30 35
 gct ctc ctg aag gac aga tcc ttc cgg ggc ccc gtc cgc cga atg agg 197
 Ala Leu Leu Lys Asp Arg Ser Phe Arg Gly Pro Val Arg Arg Met Arg
 40 45 50
 gcc tct aaa ggg gtc aag ggc aaa ggc acc gcg ctc aag ggc agg tca 245
 Ala Ser Lys Gly Val Lys Gly Lys Gly Thr Ala Leu Lys Gly Arg Ser
 55 60 65 70
 gac gct gac ctg gtg gtg ttc ctt aac aat ctc acc agc ttt gag gat 293
 Asp Ala Asp Leu Val Val Phe Leu Asn Asn Leu Thr Ser Phe Glu Asp
 75 80 85
 cag tta aac caa cag gga gtg ttg att aag gaa att aag aaa cag ctg 341
 Gln Leu Asn Gln Gln Gly Val Leu Ile Lys Glu Ile Lys Lys Gln Leu
 90 95 100

tgc gag gtt cag cat gag aga cgt tgt gga gtg aag ttt gag gtc cac 389
 Cys Glu Val Gln His Glu Arg Arg Cys Gly Val Lys Phe Glu Val His
 105 110 115

agt tta agg agt ccc aac tcc cgg gct ctg agc ttc aag ctg agc gcc 437
 Ser Leu Arg Ser Pro Asn Ser Arg Ala Leu Ser Phe Lys Leu Ser Ala
 120 125 130

ccc gac ctg ctg aag gag gtg aag ttt gat gtg ctg cca gcc tat gat 485
 Pro Asp Leu Leu Lys Glu Val Lys Phe Asp Val Leu Pro Ala Tyr Asp
 135 140 145 150

tta ctg gat cat ctt aac atc ctc aag aag cct aac caa caa ttc tac 533
 Leu Leu Asp His Leu Asn Ile Leu Lys Lys Pro Asn Gln Gln Phe Tyr
 155 160 165

gcc aat ctc atc agt ggc gta ccc gcc ggg aag gag ggc aag tta tcg 581
 Ala Asn Leu Ile Ser Gly Val Pro Ala Gly Lys Glu Gly Lys Leu Ser
 170 175 180

atc tgc ttt atg ggg ctt cag aag tac ttc ctg aac tgt cgc cca acc 629
 Ile Cys Phe Met Gly Leu Gln Lys Tyr Phe Leu Asn Cys Arg Pro Thr
 185 190 195

aag ctg aag cgc ctc atc cgc ctg gtc acg cac tgg tac caa ctg tgt 677
 Lys Leu Lys Arg Leu Ile Arg Leu Val Thr His Trp Tyr Gln Leu Cys
 200 205 210

aag gag aag ctg ggg gac ccg ctg ccc cca cag tat gcc ctg gag ctg 725
 Lys Glu Lys Leu Gly Asp Pro Leu Pro Pro Gln Tyr Ala Leu Glu Leu
 215 220 225 230

ctc aca ctc gat gcc tgg gag tat ggg agt cga gta act aaa ttc aac 773
 Leu Thr Leu Asp Ala Trp Glu Tyr Gly Ser Arg Val Thr Lys Phe Asn
 235 240 245

aca gcc cag ggc ttc tga accgtcttgg aactggtcac caagtacaaa 821
 Thr Ala Gln Gly Phe
 250

cagcttcaaaa tctactggac agtgtattat gactttcaac accaggaggt ctccaaatac 881

ctgcacagac agctcagaaa a 902

<210> 28
 <211> 251
 <212> PRT
 <213> Mus musculus

<400> 28
 Met Glu Gln Asp Leu Arg Ser Ile Pro Ala Ser Lys Leu Asp Lys Phe
 1 5 10 15

Ile Glu Asn His Leu Pro Asp Thr Ser Phe Cys Ala Asp Leu Arg Glu
 20 25 30

Val Ile Asp Ala Leu Cys Ala Leu Lys Asp Arg Ser Phe Arg Gly
 35 40 45

39/42

Pro Val Arg Arg Met Arg Ala Ser Lys Gly Val Lys Gly Lys Gly Thr
 50 55 60
 Ala Leu Lys Gly Arg Ser Asp Ala Asp Leu Val Val Phe Leu Asn Asn
 65 70 75 80
 Leu Thr Ser Phe Glu Asp Gln Leu Asn Gln Gln Gly Val Leu Ile Lys
 85 90 95
 Glu Ile Lys Lys Gln Leu Cys Glu Val Gln His Glu Arg Arg Cys Gly
 100 105 110
 Val Lys Phe Glu Val His Ser Leu Arg Ser Pro Asn Ser Arg Ala Leu
 115 120 125
 Ser Phe Lys Leu Ser Ala Pro Asp Leu Leu Lys Glu Val Lys Phe Asp
 130 135 140
 Val Leu Pro Ala Tyr Asp Leu Leu Asp His Leu Asn Ile Leu Lys Lys
 145 150 155 160
 Pro Asn Gln Gln Phe Tyr Ala Asn Leu Ile Ser Gly Val Pro Ala Gly
 165 170 175
 Lys Glu Gly Lys Leu Ser Ile Cys Phe Met Gly Leu Gln Lys Tyr Phe
 180 185 190
 Leu Asn Cys Arg Pro Thr Lys Leu Lys Arg Leu Ile Arg Leu Val Thr
 195 200 205
 His Trp Tyr Gln Leu Cys Lys Glu Lys Leu Gly Asp Pro Leu Pro Pro
 210 215 220
 Gln Tyr Ala Leu Glu Leu Leu Thr Leu Asp Ala Trp Glu Tyr Gly Ser
 225 230 235 240
 Arg Val Thr Lys Phe Asn Thr Ala Gln Gly Phe
 245 250

<210> 29
 <211> 1322
 <212> ADN
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (34)..(1128)

<400> 29
 gaggcagttc tgttgccact ctctctcctg tca atg atg gat ctc aga aat acc 54
 Met Met Asp Leu Arg Asn Thr
 1 5

cca gcc aaa tct ctg gac aag ttc att gaa gac tat ctc ttg cca gac 102
 Pro Ala Lys Ser Leu Asp Lys Phe Ile Glu Asp Tyr Leu Leu Pro Asp
 10 15 20

acg tgt ttc cgc atg caa atc aac cat gcc att gac atc atc tgt ggg	150
Thr Cys Phe Arg Met Gln Ile Asn His Ala Ile Asp Ile Ile Cys Gly	
25 30 35	
ttc ctg aag gaa agg tgc ttc cga ggt agc tcc tac cct gtg tgt gtg	198
Phe Leu Lys Glu Arg Cys Phe Arg Gly Ser Ser Tyr Pro Val Cys Val	
40 45 50 55	
tcc aag gtg gta aag ggt ggc tcc tca ggc aag ggc acc acc ctc aga	246
Ser Lys Val Val Lys Gly Gly Ser Ser Gly Lys Gly Thr Thr Leu Arg	
60 65 70	
ggc cga tct gac gct gac ctg gtt gtc ttc ctc agt cct ctc acc act	294
Gly Arg Ser Asp Ala Asp Leu Val Val Phe Leu Ser Pro Leu Thr Thr	
75 80 85	
ttt cag gat cag tta aat cgc cgg gga gag ttc atc cag gaa att agg	342
Phe Gln Asp Gln Leu Asn Arg Arg Gly Glu Phe Ile Gln Glu Ile Arg	
90 95 100	
aga cag ctg gaa gcc tgt caa aga gag aga gca ttt tcc gtg aag ttt	390
Arg Gln Leu Glu Ala Cys Gln Arg Glu Arg Ala Phe Ser Val Lys Phe	
105 110 115	
gag gtc cag gct cca cgc tgg ggc aac ccc cgt gcg ctc agc ttc gta	438
Glu Val Gln Ala Pro Arg Trp Gly Asn Pro Arg Ala Leu Ser Phe Val	
120 125 130 135	
ctg agt tcg ctc cag ctc ggg gag ggg gtg gag ttc gat gtg ctg cct	486
Leu Ser Ser Leu Gln Leu Gly Glu Gly Val Glu Phe Asp Val Leu Pro	
140 145 150	
gcc ttt gat gcc ctg ggt cag ttg act ggc agc tat aaa cct aac ccc	534
Ala Phe Asp Ala Leu Gly Gln Leu Thr Gly Ser Tyr Lys Pro Asn Pro	
155 160 165	
caa atc tat gtc aag ctc atc gag gag tgc acc gac ctg cag aaa gag	582
Gln Ile Tyr Val Lys Leu Ile Glu Glu Cys Thr Asp Leu Gln Lys Glu	
170 175 180	
ggc gag ttc tcc acc tgc ttc aca gaa cta cag aga gac ttc ctg aag	630
Gly Glu Phe Ser Thr Cys Phe Thr Glu Leu Gln Arg Asp Phe Leu Lys	
185 190 195	
cag cgc ccc acc aag ctc aag agc ctc atc cgc cta gtc aag cac tgg	678
Gln Arg Pro Thr Lys Leu Lys Ser Leu Ile Arg Leu Val Lys His Trp	
200 205 210 215	
tac caa aat tgt aag aag aag ctt ggg aag ctg cca cct cag tat gcc	726
Tyr Gln Asn Cys Lys Lys Lys Leu Gly Lys Leu Pro Pro Gln Tyr Ala	
220 225 230	
ctg gag ctc ctg acg gtc tat gct tgg gag cga ggg agc atg aaa aca	774
Leu Glu Leu Leu Thr Val Tyr Ala Trp Glu Arg Gly Ser Met Lys Thr	
235 240 245	
cat ttc aac aca gcc cag gga ttt cgg acg gtc ttg gaa tta gtc ata	822
His Phe Asn Thr Ala Gln Gly Phe Arg Thr Val Leu Glu Leu Val Ile	
250 255 260	

aac tac cag caa ctc tgc atc tac tgg aca aag tat tat gac ttt aaa 870
 Asn Tyr Gln Gln Leu Cys Ile Tyr Trp Thr Lys Tyr Tyr Asp Phe Lys
 265 270 275
 aac ccc att att gaa aag tac ctg aga agg cag ctc acg aaa ccc acg 918
 Asn Pro Ile Ile Glu Lys Tyr Leu Arg Arg Gln Leu Thr Lys Pro Thr
 280 285 290 295
 cct gtg atc ctg gac ccg gcg gac cct aca gga aac ttg ggt ggt gga 966
 Pro Val Ile Leu Asp Pro Ala Asp Pro Thr Gly Asn Leu Gly Gly Gly
 300 305 310
 gac cca aag cgt tgg agg cag ctg gca caa gag gct gag gcc tgg ctg 1014
 Asp Pro Lys Arg Trp Arg Gln Leu Ala Gln Glu Ala Glu Ala Trp Leu
 315 320 325
 aat tac cca tgc ttt aag aat tgg gat ggg tcc cca gtg agc tcc tgg 1062
 Asn Tyr Pro Cys Phe Lys Asn Trp Asp Gly Ser Pro Val Ser Ser Trp
 330 335 340
 att ctg ctg gtg aga cct cct gct tcc tcc ctg cca ttc atc cct gcc 1110
 Ile Leu Leu Val Arg Pro Pro Ala Ser Ser Leu Pro Phe Ile Pro Ala
 345 350 355
 cct ctc cat gaa gct tga gacatatagc tggagaccat tctttccaaa 1158
 Pro Leu His Glu Ala
 360 365
 gaacttacct cttgccaaag gccatttata ttcatatagt gacaggctgt gctccatatt 1218
 ttacagtcat tttggtcaca atcgagggtt tctggaattt tcacatccct tgtccagaat 1278
 tcattcccct aagagtaata ataaataatc tctaacacca aaaa 1322

<210> 30
 <211> 364
 <212> PRT
 <213> Homo sapiens

<400> 30
 Met Met Asp Leu Arg Asn Thr Pro Ala Lys Ser Leu Asp Lys Phe Ile
 1 5 10 15
 Glu Asp Tyr Leu Leu Pro Asp Thr Cys Phe Arg Met Gln Ile Asn His
 20 25 30
 Ala Ile Asp Ile Ile Cys Gly Phe Leu Lys Glu Arg Cys Phe Arg Gly
 35 40 45
 Ser Ser Tyr Pro Val Cys Val Ser Lys Val Val Lys Gly Gly Ser Ser
 50 55 60
 Gly Lys Gly Thr Thr Leu Arg Gly Arg Ser Asp Ala Asp Leu Val Val
 65 70 75 80
 Phe Leu Ser Pro Leu Thr Thr Phe Gln Asp Gln Leu Asn Arg Arg Gly
 85 90 95
 Glu Phe Ile Gln Glu Ile Arg Arg Gln Leu Glu Ala Cys Gln Arg Glu
 100 105 110

Arg Ala Phe Ser Val Lys Phe Glu Val Gln Ala Pro Arg Trp Gly Asn
 115 120 125
 Pro Arg Ala Leu Ser Phe Val Leu Ser Ser Leu Gln Leu Gly Glu Gly
 130 135 140
 Val Glu Phe Asp Val Leu Pro Ala Phe Asp Ala Leu Gly Gln Leu Thr
 145 150 155 160
 Gly Ser Tyr Lys Pro Asn Pro Gln Ile Tyr Val Lys Leu Ile Glu Glu
 165 170 175
 Cys Thr Asp Leu Gln Lys Glu Gly Glu Phe Ser Thr Cys Phe Thr Glu
 180 185 190
 Leu Gln Arg Asp Phe Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser Leu
 195 200 205
 Ile Arg Leu Val Lys His Trp Tyr Gln Asn Cys Lys Lys Lys Leu Gly
 210 215 220
 Lys Leu Pro Pro Gln Tyr Ala Leu Glu Leu Leu Thr Val Tyr Ala Trp
 225 230 235 240
 Glu Arg Gly Ser Met Lys Thr His Phe Asn Thr Ala Gln Gly Phe Arg
 245 250 255
 Thr Val Leu Glu Leu Val Ile Asn Tyr Gln Gln Leu Cys Ile Tyr Trp
 260 265 270
 Thr Lys Tyr Tyr Asp Phe Lys Asn Pro Ile Ile Glu Lys Tyr Leu Arg
 275 280 285
 Arg Gln Leu Thr Lys Pro Thr Pro Val Ile Leu Asp Pro Ala Asp Pro
 290 295 300
 Thr Gly Asn Leu Gly Gly Gly Asp Pro Lys Arg Trp Arg Gln Leu Ala
 305 310 315 320
 Gln Glu Ala Glu Ala Trp Leu Asn Tyr Pro Cys Phe Lys Asn Trp Asp
 325 330 335
 Gly Ser Pro Val Ser Ser Trp Ile Leu Leu Val Arg Pro Pro Ala Ser
 340 345 350
 Ser Leu Pro Phe Ile Pro Ala Pro Leu His Glu Ala
 355 360

<210> 31

<211> 450

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:sonde

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☒ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.